

Cite this: *Energy Environ. Sci.*,
2026, 19, 3323

Replicability challenges in redox flow cell testing: insights from a multi-institutional study

Hugh O'Connor,^a Alexander H. Quinn,^b Edward Saunders,^{cd} Aodhán Dugan,^a Thomas R. Goodwin,^b Nadia L. Farag,^c Greta Thompson,^{cd} Ameya Bondre,^e Marina Tabuyo-Martinez,^e Hannah M. Burnett,^{fg} Thomas Y. George,^h Jordan D. Sosa,^h Carlos J. Mingoies,ⁱ Peter Nockemann,^a Clare P. Grey,^c Dominic S. Wright,^c Michaël De Volder,^d Antoni Forner-Cuenca,^e Robert A. W. Dryfe,^{fg} Michael J. Aziz,^h Ana B. Jorge Sobrido,ⁱ Fikile R. Brushett^b and Josh J. Bailey^{*a}

Flow battery research is growing at pace, given the global need for longer-duration energy storage technologies. Positioned at the intersection of several scientific and engineering disciplines, flow battery studies involve significant experimental complexity that serves as a source of variability when assessing performance. Experimental errors arise from variable flow-cell assembly practices, discrepancies in electrochemical technique protocols, inhomogeneous material properties, or uncontrolled environmental conditions—all influencing the metrics reported across laboratories. Nonetheless, the magnitude of this variability in performance indicators from typical electrochemical techniques is rarely assessed. This lack of replicability testing presents challenges for interlaboratory comparison, reducing research confidence in performance ascription. We therefore performed a round-robin study involving eight participant groups (seven academic institutions) on a model flow cell system, comprising a well-studied electrolyte, in a symmetric flow-cell configuration. Despite identical cell hardware, electrolyte chemistry, and experimental prompts, appreciable differences were observed in the charge–discharge profiles, polarisation curves, and Nyquist plots resulting from participant data acquisition. The study identifies that protocol and/or in-batch material differences have clear and non-negligible effects on reported performance metrics and provides an indication of the magnitude of variabilities that can be observed for a single system. Although definitive attribution may require a larger number of participants, several plausible sources of variability were identified, and targeted follow-up testing was undertaken at the coordinating institutions to inform protocol refinement. Both electrical connections and electrolyte homogeneity in the reservoirs were observed to be non-negligible sources of variability in ohmic resistance and electrolyte utilisation, respectively. Overall, the data and insights from this well-controlled, single-electrolyte system highlight the need for greater methodological transparency, shared protocols, and standard operating procedures to reduce significant replicability error in systems of interest. Additionally, the methodology presented may guide further multi-institutional studies to address sources of variance across systems and chemistries.

Received 21st November 2025,
Accepted 7th April 2026

DOI: 10.1039/d5ee07103h

rsc.li/ees

^a School of Chemistry and Chemical Engineering, Queen's University Belfast, Belfast BT9 5AG, UK. E-mail: j.bailey@qub.ac.uk^b Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA^c Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK^d Department of Engineering, University of Cambridge, Cambridge CB3 0FS, UK^e Department of Chemical Engineering and Chemistry, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands^f Department of Chemistry, University of Manchester, Oxford Rd, Manchester M13 9PL, UK^g Henry Royce Institute, University of Manchester, Oxford Rd, Manchester M13 9PL, UK^h Harvard John A. Paulson School of Engineering and Applied Sciences, Cambridge, Massachusetts 02138, USAⁱ School of Engineering and Materials Science, Queen Mary University of London, London E1 4NS, UK

Broader context

Global net-zero emissions will require rapid expansion of renewable electricity, creating an urgent need for reliable, affordable, and scalable energy storage. Flow batteries are a promising option for long-duration storage due to their separation of power and energy, scalability, durability, and potential safety and cost advantages over lithium-ion systems. However, progress is slowed by challenges in reproducing results across laboratories, due to system complexity and varied testing practices. This international, multi-institutional collaboration provides insight into how laboratory-to-laboratory differences shape electrochemical performance metrics. We reveal how seemingly-minor inconsistencies in set-ups and protocols result in significant measurement differences, for example, a standard deviation in electrolyte utilisation of almost 10% and in area-specific resistance of up to 40% when calculated from polarisation curves. When presented with the level of parameter specification typical of published articles, participants diverged in their practices, potentially leading to significant variation in metrics derived from nominally the same system. The magnitude of these replicability errors and the lack of standard approaches highlight the urgent need for clearer testing and reporting practices. By identifying where variability arises, flagging areas of concern, and providing benchmarks for measurement error, this work supports a transition towards more replicable data and the development of next-generation flow battery materials—benefiting academia, industry, and the broader energy transition.

Introduction

Flow batteries are a promising technology for bolstering electrical grids by enabling higher penetration of intermittent renewable power sources, particularly through longer-duration energy storage,^{1,2} while also providing services such as frequency regulation and peak shaving.³ The degree to which flow batteries are successful is contingent upon system performance, reliability, life-cycle sustainability, economic value, and scalability (*i.e.*, raw material availability, robust supply chains, and manufacturability). Meeting such requirements has incentivised studies into new redox chemistries,^{4–10} and flow cell components (electrodes, membranes, flow fields) that improve performance and cost characteristics,^{11–15} leading to an expansion of the field that is evinced by the rapid growth in flow battery publications over the past two decades.¹⁶ While this burgeoning interest has considerably advanced knowledge, discrepant experimental protocols, system configurations, and data reporting have often limited qualitative and quantitative comparisons across the literature, even when employing the same electrolyte formulations or cell hardware.¹⁷ Further, practices in identifying and quantifying sources of uncertainty vary, and are absent for certain chemistries or subfields within the greater flow battery community.¹⁶

Accelerating flow battery development requires better-informed comparisons of different redox chemistries, cell components, and operating conditions, which can be enabled by knowledge of the origins and magnitudes of variability. Often, variability is expressed as repeatability, replicability, or reproducibility, but, due to the dissimilar sources of error and distinct foci of scientific disciplines, the definitions of these terms are not consistent across fields, nor necessarily within them.^{18–20} For clarity and to highlight the specific interests of this work, we adopt definitions similar to those employed by McArthur.²⁰ Here, we define repeatability of an electrochemical experiment as the measurement variability of a particular metric when multiple measurements are performed by a single team using one cell architecture.¹⁶ Replicability, which is the focus of this work, is defined here as the variability observed in such metrics when independent teams, often working in different laboratories, perform nominally identical experiments using the same cell architecture, possibly with different

auxiliary equipment. Reproducibility refers to the extent to which the same conclusions can be drawn when dissimilar cell architectures, and possibly different auxiliary equipment, are used in different laboratories. The extent to which reproducibility can be realised depends on the metric of interest and equipment employed. For example, the energy efficiency of a flow cell depends on the constituent componentry, whereas the homogeneous decay rate of a redox molecule within an electrolyte can generally (barring nuances such as electrode–electrolyte interactions) be determined independently of the flow-cell design. Because we focus on the performance of a single-cell architecture and single electrolyte chemistry, we do not discuss reproducibility in this manuscript.

Comprehensive and transparent communication of experimental details and data processing methods is critical for researchers to be able to reproduce literature studies. Experimental choices, such as component materials, pre-treatments, reagent quality, electrolyte flow rate, electrolyte tank volume, *etc.*, can influence cell performance to varying degrees, depending on the redox chemistry, device configuration, and operating conditions. Additionally, data obtained under identical conditions might be processed differently to arrive at distinct conclusions. Ultimately, concise yet detailed protocols or guidelines allow researchers to focus on the physical phenomena most relevant to their systems.^{21–24} To guide the community towards best practices in single-cell flow battery testing, it is helpful to evaluate currently employed practices and techniques. To this end, round-robin testing, also referred to as an interlaboratory comparison, quantifies the sources and extents of variations in performance metrics across research groups working on a similar problem. In electrochemistry, such studies have been employed to compare gas diffusion electrode testing platforms for fuel cells,²⁵ standardise material sets and testing protocols for proton exchange membrane water electrolyzers,²⁶ relate impedance measurements in dummy cells and 3-electrode systems,²⁷ quantify noise in corrosion measurements,²⁸ and assess variability in supercapacitor performance while identifying discrepancies in data analysis practices.²⁹ Additional round-robin studies outside of electrochemistry (*e.g.*, employing the Brunauer–Emmett–Teller theory for measuring sample surface areas³⁰ or measuring permeation through membranes³¹) can inspire strategies



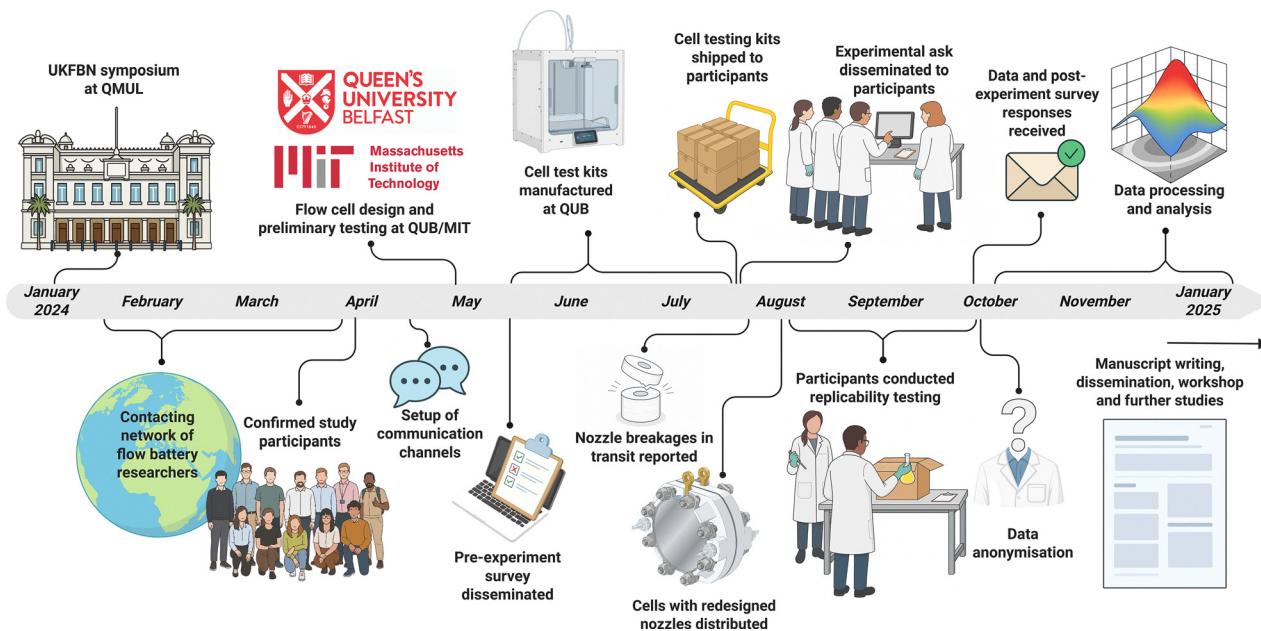


Fig. 1 A project timeline illustrating key steps taken in this multi-institutional study.

towards understanding and addressing variability (*e.g.*, exploring deviations due to analysis procedures on the same dataset³⁰ or exploring measurements of the same property across different apparatus³¹). These tests have generally demonstrated that equipment, experimental practices, and analysis methodologies contribute to discrepant results between research groups, which can be mitigated through protocol refinements. To our knowledge, no comparable study has yet been reported by the flow battery community, and such work could help elucidate the factors that contribute to variability across institutions.

This study, outlined in Fig. 1, involved participants across seven universities (eight research groups), and was borne out of discussions at the 2024 UK Flow Battery Network (UKFBN) symposium held at Queen Mary University of London (QMUL). There, attendees shared a collective perception that communication of experimental practices in the peer-reviewed literature was often insufficient, vague, or incomplete, partly due to the restricted length of published articles and a lack of readily accessible, widely accepted, and well-defined protocols for single-cell flow battery testing. Attendees also noted challenges in comparing their acquired data with those in the literature, due to variability in the apparatus used. It was generally agreed that progressing towards standards and expectations for flow battery research could enable clearer comparisons between, and more robust validations of, published work. The consensus was that such standards could support new entrants into the field, particularly those with limited background in flow-cell testing, ultimately accelerating the development of new chemistries, components, and systems. Building on the enthusiasm at the symposium and engaging several other like-minded research groups, we launched a round-robin study to focus, at least initially, on the replicability of performance metrics derived from electrochemical testing of flow cells.

Our goal was to evaluate variability in cell performance for a well-defined system, without providing overly constraining protocols. Accordingly, to quantify uncertainty in replicability testing, we used certain controls to, in principle, measure the same system in different laboratories. Specifically, we opted to compare the performance of an identical symmetric flow cell architecture using a single ferri-/ferro-cyanide-based electrolyte. To this end, the same cell kit and certain materials (membrane, electrodes, tubing, fittings, connectors) from the same supplier and batch were shipped to each participant.

Participants were asked to evaluate flow cell performance using three commonly employed electrochemical techniques: galvanostatic, or potentiostatic, step polarisation (hereafter, “polarisation”), electrochemical impedance spectroscopy (hereafter, “impedance” or “EIS”), and galvanostatic charge-discharge cycling (hereafter, “CD cycling”).

In polarisation, a current (voltage) is imposed across the cell in a sequence of discrete steps, and the corresponding cell voltage (current) is recorded at each step in an attempt to represent the steady-state response at that operating point. The resulting data are typically presented as a plot of cell voltage against current density, as illustrated in Fig. 2a. The shape of this curve provides a qualitative interpretation of performance-limiting processes, with activation losses often most evident at low current density, ohmic losses more prominent over an intermediate region, and mass transport limitations increasingly influential at higher current density. Further background on polarisation methods and their interpretation when applied to electrochemical energy systems and flow batteries are available in the literature.^{32–34}

In EIS, a small-amplitude sinusoidal current (voltage) perturbation is applied about a chosen operating point and the frequency-dependent voltage (current) response is measured to





Fig. 2 Idealised illustrations of the electrochemical measurements requested from participants. (a) Charge polarisation curve. (b) EIS Nyquist plot highlighting common spectral features and associated resistive contributions discussed in this work. (c) Galvanostatic CD cycling profile. (d) Capacity fade derived from CD cycling.

obtain the complex impedance, commonly displayed as a Nyquist plot (Fig. 2b) and/or a Bode plot. Fig. 2b is an idealised schematic of a Nyquist plot, intended to highlight common spectral features. The shape of the plot provides a qualitative separation of contributions, with the high-frequency intercept often associated with an effective cell ohmic resistance, R_{Ω} , and lower-frequency features commonly ascribed to interfacial charge transfer resistance, R_{CT} , and mass transport resistance, R_{MT} . Note that EIS is a nuanced technique whereby the measurement parameters, operating point, and the selected analysis model(s) influence interpretation. Introductions to EIS measurement and equivalent circuit analysis are provided in recent work,³⁵ along with some specific discussions of EIS applied to flow batteries.^{36,37}

In galvanostatic CD cycling, the cell is charged then discharged, repeatedly, at a controlled current between defined cut-off voltages (Fig. 2c) to determine capacity and efficiency metrics, and to track performance evolution with cycling (Fig. 2d).

In this work, the term “technique” refers specifically to the electrochemical characterisation approaches used to obtain cell data. The term “analysis” refers to the processing and quantification of variability in the resulting datasets (*e.g.*, polarisation curve fitting, Nyquist plot fitting, and CD cycling metrics). Finally, the term “methods” describes the broader experimental practices and procedures employed by participants when performing these measurements.

We observed noticeable variations in the data returned by participants, despite the use of a nominally identical cell, chemistry, and set of instructions. These differences were difficult to ascribe to a single factor but highlighted the impact of seemingly innocuous decisions in cell set-up and operation. As such, we hypothesised several factors responsible for differences and tested these hypotheses in two of the eight laboratories. These findings point to a need for greater care in reporting experimental methodologies and they encourage the establishment of general field-wide guidelines for performing specific foundational tests. Based on our findings here, we also provide some recommendations for conducting round-robin exercises and for experiment execution and reporting.

Materials and methods

Here we detail the design, execution, and materials of the round-robin study. We first discuss the surveys used to inform the study design. Subsequently, we delineate the experimental request provided to the participants and the rationale for the study decisions. We then describe the material kits shipped to each of the participating laboratories, as well as the cell componentry and electrolyte chemicals employed in the study. Finally, we detail the data handling and analysis.

Round robin design

An initial pre-experiment survey (SI “SF1 Pre-experiment survey.pdf”) was used to assess participants’: equipment and materials availability; time commitment capacity; degree of experience working with flow cells; motivation for participating in the round robin; and some typical experimental practices. This information enabled the study leads to identify a suitable chemistry and flow cell system as well as to craft a reasonable experimental ask. The key findings from this survey are detailed below (additional details are included in SI Section S1).

As expected, all participants were able to run one cell, though several indicated they could evaluate two to four cells in tandem. The survey also highlighted the diversity of electrochemical instrumentation, flow cell architectures, and balance-of-system components employed by the different participants. All participants had access to potentiostats that could achieve currents of at least 400 mA (most could achieve currents ≥ 1000 mA) and could satisfy the voltage requirements (± 0.8 V), despite an initial concern that some participants might only have access to a battery cycler rather than a potentiostat. Eight participants collected data using potentiostats from Biologic ($n = 6$), Gamry ($n = 1$), and Metrohm ($n = 1$). Among the Biologic instruments, three participants used VMP3 models (one with a VMP3B-5A booster), while others used a VSP, VSP-3e, and VSP-300 (with B-10A5V booster). The remaining participants used a Metrohm Autolab PGSTAT302N or a Gamry Interface 5000. These instruments all incorporate frequency response analysers, enabling impedance measurements. Reservoirs varied in form factor (custom glass-blown, modified burette, media bottle, centrifuge tubes) and material (glass, polypropylene).



Pumps were either diaphragm (KNF Neuberger GmbH, Germany) or peristaltic (Masterflex™, Avantor, USA; Watson-Marlow Ltd, UK; Chonry, China).

Overall, this initial pre-experiment survey qualitatively highlighted diversity and breadth in the participants' research exposure and experimental experiences, and a panoply of auxiliary equipment being routinely employed for flow battery research. These results encouraged the study leads to control for chemistry, flow cell architecture, electrode materials (and their pre-treatment), membrane materials, and certain elements of the electrochemical protocol request.

Experimental request

Based on the pre-experiment survey responses, the study leads devised the experimental request (SI "SF2 Experiment request.pdf"), detailed below, with the aim of probing replicability error across research groups, whilst also enabling the fullest participation without undue time and materials requirements. The study leads intended for the experimental request to reflect the level of detail found in 'Methods' sections of journal articles.

Our criteria for selecting a redox chemistry were: material accessibility (inexpensive and commercially available to all participants in all necessary states); safety (*e.g.*, limit use of corrosive, toxic, and/or carcinogenic chemicals); operational simplicity (low sensitivity to oxygen, minimal electrolyte processing, limited decay); and literature precedence. We also sought to avoid chemistries in which a particular group had extensive experience to avoid skewing the study results. Thus, we opted for the ferri-/ferro-cyanide ($[\text{Fe}(\text{CN})_6]^{3-}/[\text{Fe}(\text{CN})_6]^{4-}$) redox couple in near-neutral pH: an electrolyte formulation that generally met these criteria and which is known to be (electro-)chemically stable on the timescale of typical laboratory-level, single-cell CD cycling.^{38–40} A minor drawback of this model redox couple is that electrode-dependent kinetic variabilities observed with more sluggish chemistries (*e.g.*, the all-vanadium chemistry) cannot be probed. The participants were directed to pay due care to ensure their ferri-/ferro-cyanide-based electrolytes did not contact acids, to avoid the liberation of toxic HCN.

A symmetric set-up was chosen to allow for the use of a single redox couple, simplifying the experimental ask by avoiding challenges associated with selecting and operating with a second redox couple. While certain diagnostic configurations might be better-suited for performance studies (*e.g.*, single-reservoir symmetric flow cells excel at probing performance at fixed state-of-charge (SoC)⁴¹), we opted for a dual-reservoir system to enable polarisation, impedance, and CD cycling measurements in a single build (*i.e.*, without reconfiguring tubing and replacing electrolyte). As such, the symmetric system is expected to elicit some behaviours that arise in full-cell systems (*e.g.*, SoC swings), but not others (*e.g.*, crossover). Further, we did not employ volumetrically-unbalanced cells nor more detailed cycling protocols (*e.g.*, constant-current followed by constant potential cycling).⁴² While such refinements enable greater accuracy in measuring species decay,⁴² the application of equal-volume reservoirs and galvanostatic CD cycling

protocols minimised experimental complexity and provided the level of resolution needed to compare results.

Participants were asked to collect polarisation, impedance, and CD cycling data for a symmetric flow cell system consisting of two reservoirs (each containing 100 mL of aqueous electrolyte composed of 100 mM potassium ferricyanide, 100 mM potassium ferrocyanide, and 1000 mM potassium chloride) and identical 3D-printed flow cells with 16 cm² active area, inspired by a previous design described in detail below and in SI Section S1.^{16,43} Participants cut electrodes (SGL Carbon 4.65 EA) to size using their own tools (*i.e.*, scalpels, razor blades, scissors, dies). Some of those that used scalpels or razor blades additionally employed a pre-cut template to guide the cutting edge.

Polarisation data were collected using chronoamperometry or chronopotentiometry, at 50% SoC, without *iR*-compensation, over a minimum absolute current density range of 0 to 62.5 mA cm⁻², limited to a maximum absolute cell voltage of 0.8 V, and at a minimum resolution of one data point *per* second (leaving step magnitudes, durations, and ordering up to the participant). Impedance was collected at a participant-selected perturbation amplitude over a frequency range of 200 kHz to 10 mHz, about open-circuit voltage (OCV), at 50% SoC, and with 6 points of log-spaced data collected *per* decade. Impedance and polarisation data were collected at flow rates of 10, 30, and 50 mL min⁻¹. Data were additionally requested at 1 mL min⁻¹, but an oversight by the study leads in assuming that all participants' pumps could accurately deliver this flow rate resulted in an incomplete dataset. As a result, only three of the eight participants were able to acquire data at 1 mL min⁻¹. Accordingly, these data are omitted from the main text but can be found in the SI. CD cycling was conducted at a flow rate of 50 mL min⁻¹, a current density of 25 mA cm⁻², and cell cut-off voltages of ±0.5 V for a minimum of 3 days, at a minimum data resolution of 10 *s per* point.

We discouraged information sharing across institutions to elicit variabilities due to distinct choices beyond those specified, simulating the typical research environment whereby the researchers involved do not have direct real-time contact with others undertaking the same study. Participants were asked to direct questions to the study leads *via* a private communications channel (Slack Technologies LLC, USA). If queries arose that required communication to all participants, the question and response were posted in a study-wide communication channel by the leads without any information about the participant(s) who submitted the original inquiry. Two post-test surveys were included with the experimental request to collect information about specific experimental procedures, data analysis, and data reporting practices. These survey questions are included in full in the SI "SF3 Post-experiment practices survey.pdf" and "SF4 Post-experiment data analysis and reporting survey.pdf".

Materials distributed to participants

To isolate variation due to experimental practices from that arising from distinct flow cell apparatus across the



participating laboratories (thus probing replicability rather than reproducibility), we distributed kits containing identical flow cells, all internal cell components, and supporting cell hardware. The flow cell was selected because it could be manufactured and modified by one of the lead authors who designed the device. In addition, the components are relatively inexpensive and could be fabricated in bulk to send to many participants. Finally, the cell design (refined for this study in ways not expected to influence electrochemical performance) has been demonstrated to be repeatable in prior published work.⁴³ The individual components of the test kit including the flow cell are shown in SI Fig. S1. Key print settings (Table S2), additional information on the slicing and computer-aided design (CAD) software, and engineering drawings of all non-3D-printed components (Fig. S2a–c) are provided in SI Section S1. The associated print files (.gcode and .3mf) are also available in the SI “SF5 3D-printing files.zip”. Half-cell body, end plate, isolation plate, and current collector CAD files for the cell used in this study have been made available in the SI “SF6 CAD files.zip”. Finally, the guide, which specified the correct order of component assembly and provided both the torque value (5 Nm) and tightening pattern to be applied to the bolts, is available as SI “SF7 Assembly guide.pdf”.

Several concerns motivated the decision to provide electrode and membrane materials: batch-to-batch variability in electrode and membrane quality, historical changes in the materials manufacturing (both of which might vary with geography and/or vendor choice), and differing conditions for storing electrodes (*e.g.*, humidity, exposure to air, duration on shelf). Additionally, we specified for participants to not pre-treat their electrodes prior to use to avoid performance differences that may arise from in-house activation procedures that vary in protocol or equipment used. Graphite felt was sourced from a single large batch and provided as sheets ($\sim 10 \times 10 \text{ cm}^2$), from which participants cut individual electrodes for testing. Likewise, membrane samples were prepared from a single roll and supplied in individual vials of deionised water (18.2 M Ω cm). To enable reliable cell assembly and sealing, laser-cut expanded ethylene propylene diene monomer (EPDM) gaskets (including spares), and O-rings were included, along with aluminium end plates, bolts, and nuts for clamping. For integration with external circuitry, each kit also included graphite–polymer composite (PV15, SGL Carbon) and copper current collectors, and electrical clips. Fluidic components—including tubing and connectors—were provided to ensure compatibility with a wide range of laboratory set-ups. Additional tubing and clips were included to accommodate variations across systems. A bill of materials is provided in SI Table S1, listing supplier information, technical specifications, and material costs. At the time of this study, the total raw material cost *per* kit was approximately USD \$140.

Electrolyte chemicals

While the same electrolyte composition was used, participants utilised component chemicals of different purity and source. Potassium ferricyanide was sourced from Sigma Aldrich

($\geq 99.0\%$, four participants), Thermo Scientific Chemicals (99%+, for analysis, two participants; 99%+ ACS Reagent, one participant), or Fluorochem (unspecified purity/grade, one participant). Potassium hexacyanoferrate(II) trihydrate was sourced from Sigma (ACS reagent, 98.5–102.0%, one participant; ACS reagent $\geq 98.5\%$, two participants), Thomas Scientific Chemicals (ReagentPlus[®], $\geq 98.5\%$, one participant; analysis grade, 99%+, two participants), or Alfa Aesar (98%+, two participants). Potassium chloride was sourced from Sigma (ACS reagent, 99.0–100.5%, one participant; anhydrous, free-flowing, Redi-Dri[™], ACS reagent, $\geq 99\%$, one participant), VWR Chemicals ($\geq 99\%$, GPR RECTAPUR[®], one participant), Thermo Fisher Scientific (ACS Reagent, 99%+, one participant; $\geq 99.0\%$ to $\leq 100.5\%$, one participant; $\geq 99\%$, one participant), Alfa Aesar (99%, one participant), Avantor (ACS reagent, Avantor–J. T. Baker, Baker analysed; one participant). Deionised water was obtained from a MilliQ system (from which five participants reported a resistivity reading of 18.2 M Ω cm, one reported 10 M Ω cm, and one reported 4.76 M Ω cm) or a Sartorius ARIUM mini (18.2 M Ω cm, one participant). All chemicals were used as received, but shelf life was not monitored, and the electrolyte preparation procedure was left to the discretion of each researcher.

Data analysis

Data were first anonymised by JJB to minimise participant incentives to re-collect data that did not appear like the others, ensure unbiased data analyses by the study leads, and remove any potential or perceived potential for loss of reputation. The data were then analysed by a subset of the authors (AHQ, HOC, and ES) who processed all data according to prescribed protocols detailed below.

Polarisation data were aggregated into a single Excel spreadsheet with 5 columns containing current, voltage, time, flow rate, and “step” information from text files provided to AHQ. The step column was populated with a unique non-zero integer for each chronoamperometric or chronopotentiometric step. Otherwise, this field was assigned a zero (indicating a portion of the data not analysed for the polarisation curve). Polarisation curves were obtained by averaging the current and potential data for the last 10 s within each step to limit inclusion of transient behaviour. Steps shorter than 10 s were excluded from the analysed data (*e.g.*, due to a voltage limit being achieved before 10 s of the step had elapsed).

Impedance data were aggregated into a second spreadsheet with 4 columns containing $\text{Re}(Z)$, $-\text{Im}(Z)$, frequency, and flow rate data by AHQ and then processed by ES. The ambiguous nature of interpreting impedance data has a large impact on the resulting quantification of physical phenomena. To mitigate this, all impedance data were analysed by the same author (ES) using the same software, same equivalent circuit model, and the same fitting procedure. Data processing and fitting were performed in Python using `impedance.py`.⁴⁴ Data at each flow rate were fit to a modified Randles equivalent circuit model that has previously been used to describe flow cells.^{14,37,45,46} This circuit was chosen for its ability to balance



few fitting parameters, while still capturing the major physical phenomena of a flow cell with its constituent circuit elements.⁴⁷ The model, see Fig. 2b for reference, incorporates inductance (L , H) of the electrical leads, the combined ohmic resistance of the membrane and other ionic/electronic conducting components (R_{Ω} , Ω cm²), charge-transfer at the electrode–electrolyte interface (R_{CT} , Ω cm²), mass transfer to the electrode surface (finite length Warburg element, W , Ω cm²), and a constant phase element (CPE, Ω cm²) associated with double-layer capacitance and spatially-dependent behaviour (e.g., heterogeneous reactive/capacitive behaviours across electrode surfaces⁴⁸). In brief, the fitting involved minimising the sum of the squares of the unweighted residuals (between the model and data) using the Levenberg–Marquardt algorithm.

CD cycling data were compiled into a third spreadsheet and processed by HOC. The raw voltage–time data were segmented into individual charge and discharge cycles. For each cycle, the charge and discharge steps were extracted and assigned two columns each containing time and voltage. Average charge and discharge voltages were calculated *per* half-cycle by averaging across all potential data in each column. Coulombic efficiency (CE) was determined by dividing the discharge capacity by the charge capacity of the prior step. Electrolyte utilisation (EU) was calculated as the quotient of actual discharge capacity and theoretical capacity (Eqn S1 in Section S2). These metrics were computed for cycles 2 through 20 of each dataset, resulting in 19 evaluated cycles *per* test. Cycle 1 was excluded from analysis to minimise initial conditioning effects, such as electrode wetting and membrane “break-in”, on the cell metrics. The mean and standard deviation for both CE and EU were then calculated across these 19 cycles to assess performance and repeatability within each dataset.

SI “SF8 Data and code.zip” contains data and analysis code. This includes the polarisation, impedance, and CD cycling data spreadsheets. We also provide the MATLAB code used to analyse the polarisation, impedance, and CD cycling data. The fitting routine for the impedance is also provided as a Python script. Fit parameters and goodness-of-fit metrics (standard deviations) for each fit are provided in tables included in “SF8 Data and code.zip”. Additional fitting details are provided in Section S2. In this work, data collected by the same participants are plotted in the same colour. We purposely omit any information that would allow the other participants or the readership to connect a particular participant to a specific dataset.

Eight datasets are anonymised and labelled “P1, P2, P3 . . .”. Five datasets, tagged with a “b” (P2b, P3b, P6b, P7b, and P8b), were partially or fully re-collected due to issues identified during post-test processing that fell beyond the remit of replicability error (*i.e.*, precipitation (P2), incorrect electrolyte composition (P3, P6), a leak during CD cycling (P7), or a change to conditioning procedure due to low accessed capacity (electrolyte pumped overnight through the cell) in addition to performing the experiment in the ambient environment as opposed to inside a glovebox (P8)). Here, we focus on the datasets free from these issues (P1, P2, P3b, P4, P5, P6b, P7, and P8b) for

polarisation and impedance and (P1, P2b, P3b, P4, P5, P6b, P7b, and P8b) for CD cycling and correlations between the different experiments. For completeness, the original and re-collected datasets are compared in SI Section S2 (Fig. S3–S5). Also note that P1 deviated from the instructions to use a two-reservoir configuration for all experiments. Rather, P1 employed a single-reservoir cell for their polarisation and impedance experiments and then switched to a two-reservoir system for CD cycling after evacuating the cell/reservoir of electrolyte and replacing it. We elected to keep these data to, where appropriate in this text, highlight the effect of electrolyte exchange and to compare single-electrolyte reservoir polarisation behaviour with that obtained with a two-reservoir system.

Results and discussion

Our goal was to measure variability in polarisation, impedance, and CD cycling for a relatively simple flow cell system (a symmetric flow cell, two equal-volume electrolyte reservoirs, and a ferri-/ferro-cyanide-based electrolyte) given the same flow cell equipment as well as the same assembly and operation instructions. Prior to sharing the experimental request, AHQ and HOC tested identical flow cell configurations under the same conditions in their respective labs to refine experimental conditions and demonstrate that the flow cell system was reasonably replicable (SI Fig. S6). While we specified the experimental ask with sufficient information to perform the experiments, participants still needed to make set-up and operational choices for unspecified portions of the procedures. These distinct choices in experimental practices, summarised in Fig. 3, included material preparation, cell connections, ambient temperature, electrochemical protocol, order of application of electrochemical techniques, and system operation (*e.g.*, electrolyte stirring, sparging, and pump calibration). Further differences in participants’ set-ups are described in SI Tables S3–S5.

We recognise that experimental results can be influenced by the level of experience of the participant and the lab, as well as access to flow battery resources. This study lowers the barrier-to-entry by supplying flow cell components and choosing an affordable electrolyte system, providing a proxy for the minimum variability expected among researchers with a range of experience. Additionally, we quantified participant experience by surveying the length of time each researcher had been working with flow batteries. Based on this self-reported measure, no clear correlation was observed between participant experience and any of the CD cycling, polarisation, or impedance metrics analysed in this study. However, we note that this metric provides only an imperfect and indirect indication of practical experience. Recommendations of best practices can further decrease variability between experience levels by clarifying flow battery protocol for new researchers.

As we explore in the next sections, differences in these set-ups, not necessarily limited to those documented, affect the data acquired from the application of polarisation, impedance,



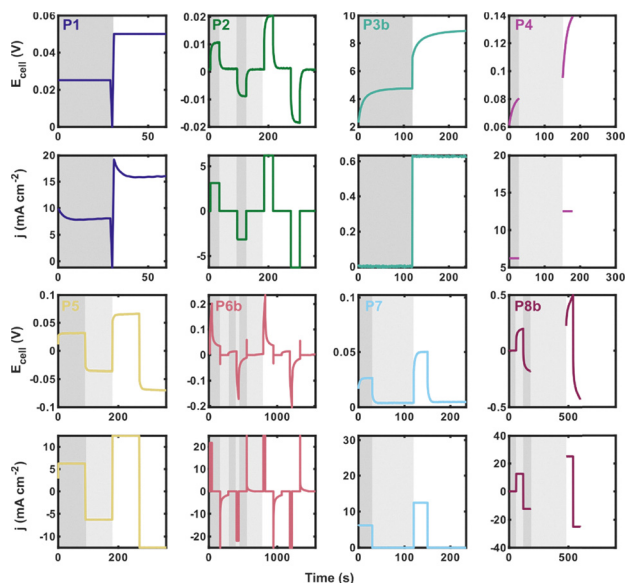


Fig. 4 Cell potential (E_{cell}) and current density (j) vs. time *per* participant, showing representative data of the various protocols used to collect polarisation data. Data are collected using repeating units consisting of single or several techniques (e.g., constant current, constant potential, open circuit), where each technique may employ different parameters (step duration, potential, or current magnitude). Further, the technique ordering varies by participant. Each panel contains two repeating units, where the first repeating unit is highlighted. Different shades of grey are used to indicate a change in technique or a change in parameter (e.g., a potential step). Data collected here are for identical flow cell builds (same components and material origins) in a symmetric two-reservoir configuration. Each reservoir contained 100 mL of 100 mM $\text{K}_3\text{Fe}(\text{CN})_6$, 100 mM $\text{K}_4\text{Fe}(\text{CN})_6$, and 1 M KCl. Note that P1 is an exception, in that a single-reservoir configuration was used. Electrolyte flow rate was 50 mL min^{-1} .

compared to *ca.* 220 mV when uncorrected. Further, upon *iR*-correction, the coefficient of variation increases for the all-participant dataset (44% to 80%) but decreases for the excluding dataset (27% to 13%) (Fig. 5f). This implies that differences in ohmic resistance account for a major portion of the observed standard deviation in the excluding dataset, but not for all participants. Data obtained at other flow rates exhibit similar increasing standard deviations (and similar values across flow rates) and constant coefficients of variation across current density (but variable with flow rate), all of which is shown in SI Fig. S8 for all participants.

Several potential contributions to the above variability can be inferred from the data and from post-experiment survey information. The notable increase in coefficient of variation near 0 mA cm^{-2} seemingly reflects non-zero OCV values (ranging across $[-3, 28] \text{ mV}$ at 50 mL min^{-1}) which likely arise from deviations in the system SoC from 50%. This could be a consequence of prior polarisations, given that prior-collected polarisation data show OCVs closer to zero (ranging across $[-8, 12] \text{ mV}$ at 10 mL min^{-1} and $[-4, 18] \text{ mV}$ at 30 mL min^{-1}). Given that most participants collected data in order of increasing flow rate, this reflects a general shift towards more positive OCVs consistent with the positively biased polarisations. While the polarisation protocol is possibly responsible for some

differences, a comparison of the P1 and P2 datasets shows that nearly identical results are obtainable with two distinct polarisation protocols. This is not believed to be entirely coincidental: although P1 employed a single electrolyte reservoir and P2 employed two reservoirs (on average maintaining SoC with positive and negative polarisation), the composition of the electrolyte entering into each half-cell at each potential/current density should, on average, be the same. That the ohmic losses are the same might suggest these two cells are not compromised by contact resistances due to the use of 4-probe connections (SI Table S3, *vide infra*). This suggests that it is possible, under the right conditions, to compare measurements between single- and dual-reservoir configurations and reflects nuance in defining replicable polarisation protocols. In the *iR*-corrected data at 25 mA cm^{-2} , 6/8 participants are within 26 mV of each other suggesting that P4 and P8b might be considered outliers. Interestingly, the P3b dataset agrees with this group at low current densities but begins to diverge at *ca.* 40 mA cm^{-2} . We posit this arises from the choice to polarise only positively in a two-reservoir setup using a protocol which employs longer pulses (120 s) than most other participants (SI Table S4), ultimately leading to SoC drift during the polarisation experiment. P7, while only positively polarising, only collected up to 25 mA cm^{-2} before the SoC shifted enough to skew the results (but is systematically excluded from the excluding dataset for consistency). P4 employs a similar protocol to P3b but reported exposure of their brass current collector to the electrolyte (SI Table S5). In this case, unfavourable side redox reactions might influence the potential through corrosion. Additionally, this compromised interface between the current collector and graphite-polymer composite may explain the higher resistance of P4. No clear explanations were found for the performance deviation of P8b.

Several protocol refinements can reduce issues due to SoC drift in polarisation measurements. The employment of a single, instead of dual, reservoir system eliminates SoC drift (assuming minimal species decay and faradaic side reactions). However, this diagnostic configuration is not applicable to CD cycling studies and does not represent flow battery systems in the field (which require at least two reservoirs). For symmetric or full-cell systems, polarisation techniques can be designed to, on average, correct for SoC (*i.e.*, by positive and negative polarisation). The change in SoC (Δx , —) can be estimated using eqn (1), where j (mA cm^{-2}) is the current density, A (cm^2) is the cell geometric area, t (s) is the duration of the constant-current (or constant-potential) step, V (L) is the electrolyte volume of a single reservoir, n ($\text{mol}_e \text{ mol}^{-1}$) is the moles of electrons transferred *per* mole of reacting redox species (here, $n = 1$), F ($96485 \text{ C mol}_e^{-1}$) is the Faraday constant, and C (mol L^{-1}) is the total concentration of the active species.

$$\Delta x = A \int_0^t j dt / VnFC \quad (1)$$

This suggests that shorter-duration polarisation steps (smaller t) lead to reduced SoC drift. However, the duration of the



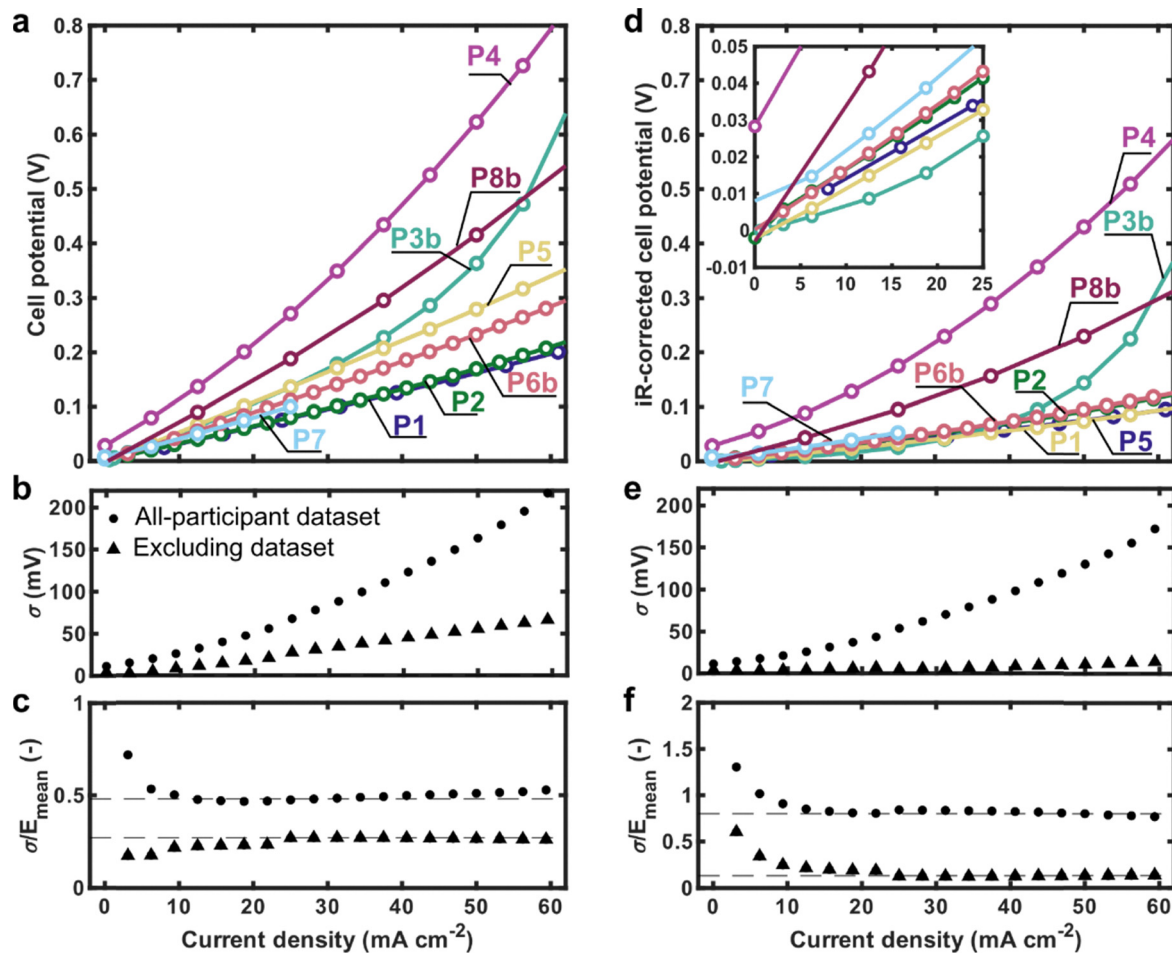


Fig. 5 Cell polarisation data at 50 mL min^{-1} , including metrics of spread. (a) Uncorrected cell potential, and corresponding (b) standard deviation and (c) coefficient of variation. (d) iR -corrected cell potential, assuming x -axis intercept in the Nyquist plot corresponds to ohmic contribution, and corresponding (e) standard deviation and (f) coefficient of variation. In (b), (c), (e), and (f) two groupings of data are presented: one including all participants ($N = 8$ participants, 1 cell *per* participant) and a subset ($N = 4$ participants, 1 cell *per* participant) which excludes P3b, P4, P7, and P8b. Horizontal dashed lines in (c) and (f) highlight the near-constant coefficient of variation for each dataset.

polarisation step must be sufficiently long to avoid capturing transient effects in the averaged portion of the potential or current data. Otherwise, the polarisation curve will likely over-predict the steady-state performance (lower potential or higher current density than observed at steady-state). The residence time (τ , s) of electrolyte passing through the electrode is a convenient means to approximate a transience timescale. Eqn (2) provides an estimate of the scale of residence time in the porous electrode, where V_{el} (m^3) is the total electrode volume (including pore and solid volume), ε (—) is the electrode porosity, and Q ($\text{m}^3 \text{ s}^{-1}$) is the volumetric flow rate of the electrolyte through the electrode.

$$\tau = V_{el}\varepsilon/Q \quad (2)$$

Increased flow rates decrease the residence time allowing for shorter pulses. However, the flow rate can also increase the current density (and thus SoC drift) at a constant potential, depending on the system. A solution to this trade-off is larger

electrolyte volumes to decrease SoC drift throughout each polarisation step and/or enable longer-duration pulses.

Impedance

Electrochemical impedance spectroscopy (EIS) measurements allow the probing of resistive losses within the flow cell, associated with various timescales and spatial locations.⁴⁹ Ideally, the flow cell impedance data are fit to a well-defined equivalent circuit model that separates the resistive contributions into ohmic (a series combination of electronic, ionic, and contact resistances), charge transfer, and mass transport resistances. Such experiments thus provide complementary information to the polarisation analysis, while also allowing for a measure of ohmic resistance useful for performing iR corrections. We present impedance data represented as Nyquist plots, together with traces associated with fits to an equivalent circuit model at three different flow rates (Fig. 6a–c). The equivalent circuit and resistance breakdowns are shown in Fig. 6d. Data collected at 1 mL min^{-1} can be found in SI Fig. S9. The





Fig. 6 Fitted impedance data for each participant at several different flow rates. Nyquist plots at (a) 10, (b) 30, and (c) 50 mL min⁻¹. The data fit ranges from 1 MHz to 1 mHz. Depending on the dataset, this involves some extrapolation at lower and/or higher frequencies. High-frequency imaginary components are truncated if lower than $-5 \Omega \text{ cm}^2$ to avoid overlapping plots. In (a)–(c), the grey dashed lines indicate the x-axis ($y = 0$). (d) Resistance breakdown corresponding to each Nyquist plot and equivalent circuit model used for all fits. $N = 1$ cell per participant.

impedance data presented here were often collected either directly prior to, or directly after, the polarisation curves.

While most of the datasets fit well using the same equivalent circuit model, the feature forms in the Nyquist plots vary substantially. Specifically, Fig. 6a–c illustrates that the position of the x-axis intercept as well as the size and shape of the higher-frequency charge transfer and lower-frequency mass transfer arcs all vary considerably across the study. Consequently, variable resistances (R_{Ω} , R_{CT} , and R_{MT}) are extracted from the data (Fig. 6d). Using the excluding dataset employed in polarisation, R_{Ω} ($2.22 \pm 0.84 \Omega \text{ cm}^2$) is the largest contribution to resistance and uncertainty in the ASR_{EIS} , followed by R_{MT} ($1.29 \pm 0.31 \Omega \text{ cm}^2$), and R_{CT} ($0.47 \pm 0.34 \Omega \text{ cm}^2$). Further, the high-frequency features below the x-axis, attributed to an ideal inductor, vary in length. This may arise from the arrangement and configuration of electrical leads and can influence the fit and interpretation of the higher-frequency data corresponding to the ohmic (R_{Ω}) and kinetic resistances (R_{CT}). Limited information about electrical configurations, namely connectors and use of a 2- or 4-probe connection, are included in Table S3. 4-point connections can be used to mitigate lead and lead-connection resistances. Indeed, the participant group using 4-point connections (P1, P2, P6b) observed generally a tighter distribution about a lower mean for R_{Ω} ($1.85 \pm 0.45 \Omega \text{ cm}^2$ for 50 mL min⁻¹).

Nevertheless, it is worth reflecting on the impedance behaviour of cells operated by different users, as a function of flow rate. As expected, for most participants, R_{Ω} remains nearly constant as flow rate is increased (10, 30, and 50 mL min⁻¹). The two exceptions to this are a moderate decrease upon increasing the flow rate for P3b, and a relatively large decrease upon increasing flow rate for P7. Indeed, for P7, there is evidence for an evolving state of the flow cell in general across the experiment, as suggested by the atypically large impedance at 10 mL min⁻¹. Although difficult to determine here, such differences in system evolution might arise from combinations of material preparation, cell operation procedure, and/or measurement timing (e.g., soaking membranes in electrolyte for an unspecified duration, or variable cell conditioning steps, Table S5). R_{CT} , like R_{Ω} , remains almost invariable upon change in flow rate, except again for P3b and P7. In some cases, R_{CT} cannot easily be resolved (e.g., fits for R_{CT} of P2 approach zero). Further, the error in fitting R_{CT} is often comparable to its absolute value and therefore reduces confidence in its value (see “SF8 Data and code.zip”). Notably, a trend of decreasing R_{MT} upon increasing flow rate is observed for most participants, agreeing with the expectation of improved mass transport rates of active species to and from the electrode surfaces. However, the magnitude of the reduction varies, with relatively small changes observed for P8b, for example.



Although there are significant differences in magnitude across institutions for all three resistance types, there are cases (e.g., P1 and P2), where the datasets have similar total ASRs across the three flow rates (and similar polarisation behaviour at 50 mL min^{-1} , Fig. 5). However, upon closer inspection, they have different plot structures and different resistance breakdowns, suggesting challenges in unambiguously attributing the Nyquist plot features to specific phenomena.

Charge–discharge cycling

Symmetric cell CD cycling can help assess durational performance of a particular redox species, electrolyte formulation, or cell component.⁵⁰ For simplicity, we elected to use a capacity-balanced configuration, whereas a capacity-imbalanced configuration is often advantageous for measuring slow decay processes. A representative cycle of each participant's cycling data (cycle 10) is shown in Fig. 7a. While all the traces have a similar shape, variations in charge–discharge voltages and accessible capacities are evident. In most datasets, (dis)charging begins between ± 0 and $\pm 0.05 \text{ V}$, corresponding to approximately 0% SoC. An exception to this trend is P8b, which exhibits an abrupt increase from 0 V to around $\pm 0.12 \text{ V}$ shortly after (dis)charging commences. This behaviour is consistent with the more resistive polarisation response observed for P8b; however, this is not reflected in its impedance response, which is comparable to those of other participants. The corresponding iR -corrected cycling data (cycle 10) are shown in Fig. 7b. These data are generally in greater agreement than those which are not iR -corrected, with greater overlap (besides P8b), at least until they diverge due to the different EUs. The absolute rise in potential at the end of each charge–discharge step is relatively large when compared to the potential itself and thus it is not thought that iR correction will affect EU greatly.

The corresponding performance metrics (EU, CE, and decay rate), calculated for each cycle and averaged across all 19 CD cycles, are shown in Fig. 8a–c. Note that the error bars reflect the standard deviation across the 19 cycles. EU (Fig. 8a) exhibited the largest relative variability, with an average of $87.5 \pm 9.3\%$. Elevated onset voltages for P8b may have contributed in part to the lowest EU of all datasets by an early attainment of the voltage limits. Another outlier was P1, which exhibited an average discharge time of 5018 s across 19 CD cycles, exceeding the maximum theoretical charge/discharge time of 4824 s, thus resulting in an $\text{EU} > 100\%$. Review of the post-experiment survey revealed that P1 changed from a single to dual electrolyte tank configuration without fully evacuating residual electrolyte. This additional and unaccounted for electrolyte volume resulted in an unphysical capacity. Excluding the two outlying datasets (P1 and P8b) reduced the standard deviation to $\pm 6.3\%$, which, while improved, remains higher than that observed for other performance metrics calculated from CD cycling. The CE and capacity fade rate for P1 were otherwise comparable to those of other datasets. Cases such as P1 highlight the importance of detailed and clearly defined testing protocols for flow cell studies, to prevent the misreporting of metrics such as EU.



Fig. 7 Representative charge and discharge curves (cycle 10) from each participants' CD cycling data for the eight participants ($N = 8$ participants, 1 cell per participant) (a) non- iR -corrected data. (b) iR -corrected data, where t_{theo} is the theoretical duration of a single charge or discharge (4824 s).

Although no single reason emerged to describe this variation in EU, we can speculate on possible explanations based on participant responses to the post-experiment survey. These include bypass of fluid from flow cell outlet to inlet stream due to non-ideal placement of tubing in reservoirs (especially for cells without tank mixing), loss of active materials (e.g., accumulation of ferri-/ferro-cyanide on reservoir walls due to agitation and splashing, which may remove active species from circulation), unintended reaction with cell components (e.g., brass), variable electrode performance (e.g., due to differences in flow rate, electrode wetting extent, temperature), and variable ohmic resistance. Additionally, smaller effects from the variation in measurement of electrolyte concentrations, electrolyte volume, cross-over behaviour, amongst other unidentified possibilities, might compound differences. Across 19



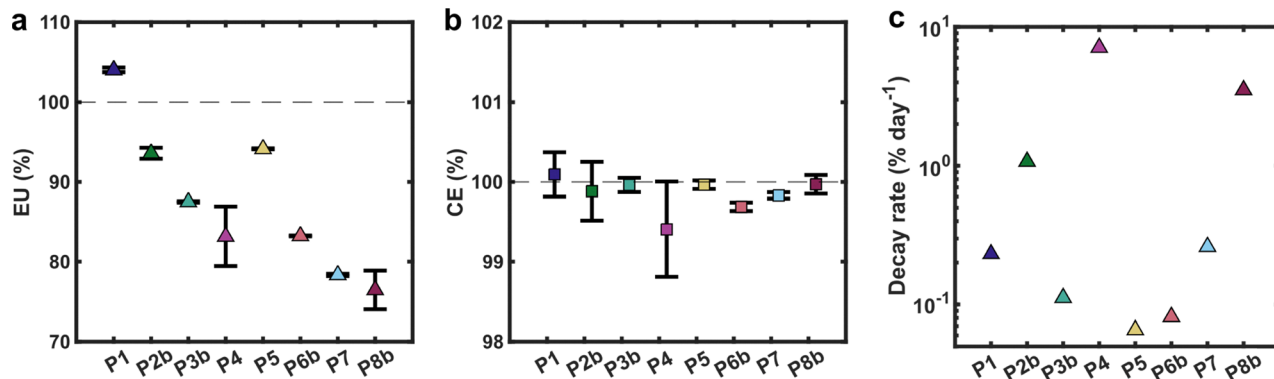


Fig. 8 (a) Electrolyte utilisation (EU) and (b) coulombic efficiency (CE) averaged across all cycles. Error bars represent the standard deviation in the metric per cycle ($N = 19$ cycles, 1 cell per participant). (c) Linearly fit time-denominated capacity fade.

cycles for all participants, the average CE was $99.8 \pm 0.2\%$ (Fig. 8b). Ideally, given the stability of the ferri-/ferro-cyanide couple and the symmetric cell build, CE should be near 100%, with values below 100% indicating capacity loss. Apparent decay rates, linearly fit to the discharge capacity over the 19 cycles, span from ~ 0.05 to $\sim 8\%$ day⁻¹ (Fig. 8c). While contact of the electrolyte with the current collector might account for the accelerated decay in P4, a substantial spread in decay rate is observed where such acute issues were not reported by the other participants (Table S5). The lowest observed decay rates are higher than those reported in arguably the most similar system: apparent decay rates within a symmetric cell system (involving a volumetrically-imbalanced set-up and constant voltage cycling) employing 100 mM ferricyanide, 100 mM ferrocyanide, and no supporting salt, suggests an immeasurable decay at pH 7, and $< 0.01\%$ day⁻¹ at pH 12 (whereas our electrolyte was measured at *ca.* pH 10).⁴⁰ Other symmetric cell investigations have reported values of *ca.* 0.0068% day⁻¹.³⁸ Decay rates measured in literature in full-cell systems which employ ferri-/ferro-cyanide in one of the electrolytes at neutral pH are also lower (*e.g.*, 0.014% day⁻¹, 0.00027% day⁻¹).^{51,52} Differences in electrolyte compositions, active species concentrations, cell configurations, and electrochemical protocols used across the literature to measure these apparent decay

rates challenge direct comparison. For instance, because we employed a capacity-balanced system (equal volume and initial electrolyte composition in each reservoir), the apparent decay rates might reflect charge imbalance between half-cells. Further, visible light, particularly of shorter wavelength (< 500 nm), has been reported to decompose ferri-/ferrocyanide.^{53–59} Variable lighting configurations, due to differences in laboratory lighting and/or exposure to sunlight, may affect the decay rate. Finally, oxygen ingress into the system might oxidise ferrocyanide.

Averaging over 19 cycles obscures some of the nuanced cycle-to-cycle differences, as shown in Fig. 9. In Fig. 9a, the absolute values of the average charge and discharge voltages show steady or slightly-increasing behaviour for some cells (P1, P2b, P3b, P5, and P7) and settling behaviour for others (P4, P6b, and P8b). Such time-dependent changes in system state impact the means and standard deviations reported in Fig. 8. Differences between charge and discharge voltages (Fig. 9a) result in voltage efficiencies which deviate from 100%, which would be expected in an ideal symmetric cell. Systematic positive or negative deviations here are indicative of an imbalance between either the performance of two half-cells and/or the electrolytes in the tanks. The CEs from multiple participants *per* cycle are often indistinguishable from 100%, possibly due to the

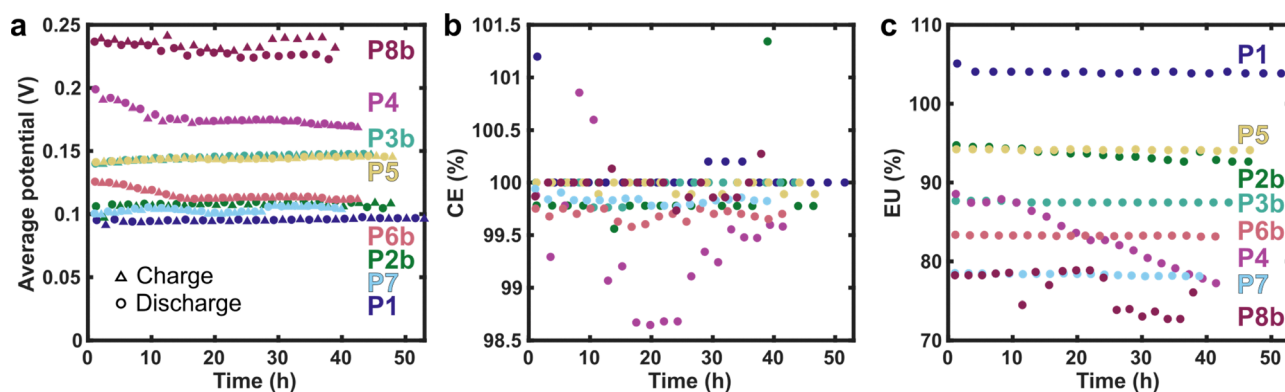


Fig. 9 Per cycle variations in symmetric flow cell metrics. (a) Absolute value of the average charge voltage and discharge voltage *per* cycle vs. time for all 19 cycles. (b) Coulombic efficiency *per* cycle vs. time for all 19 cycles. (c) Electrolyte utilisation *per* cycle vs. time for all 19 cycles, 1 cell *per* participant.



time-resolution at which the CD cycling data are collected (1 s frequency, Fig. 9b); however, CEs of several participants are consistently below 100%. EU (Fig. 9c) is approximately constant for most participants but varies substantially between participants. P4 and P8b both initially exhibited stable EU values that began to decline from cycles 5 and 11, respectively, suggesting an event that initiated subsequent decay. For P4, a crack in the cell body was reported, accompanied by black deposits on the brass current collector and a small amount of crystallisation in the electrolyte tanks. In contrast, P8b did not report any clear cause; however, the observed variations in capacity may reflect side reactions or gradual changes in cell performance over time, potentially influenced by factors such as temperature fluctuations. P7 reported a minor loss of active material prior to cycling, which explains the lower capacity yet comparable average potential to the other participants (Table S5).

Correlations between polarisation, impedance, and CD cycling

Ideally, experimental measurements such as polarisation or impedance can predict aspects of flow cell performance more rapidly than CD cycling. However, the extent to which these individual techniques predict each other's performance depends on multiple factors, even under seemingly well-defined conditions (*i.e.*, identical cells and operating parameters). System characteristics, including the redox chemistry, operating conditions, system history, and cell architecture, can all influence the correlation of impedance, polarisation, and CD cycling. As such, we sought to correlate the three techniques we studied. In Fig. 10, three correlations are shown to reflect connections between polarisation, impedance, and CD cycling in our system. Fig. 10a shows that the ASR derived from impedance (ASR_{EIS}) correlates with the ASR calculated from the polarisation curve slope (ASR_{pol} , estimated by regressing a line to the polarisation curves between 0 and 25 mA cm⁻²). However, depending on the flow rate and the participant, the extent of deviation between these two values shows that impedance is an imperfect predictor of ASR_{pol} : measurements at

higher flow rate generally result in a better correlation between ASR_{pol} and ASR_{EIS} although this might be a consequence of experiment ordering.

ASR_{pol} correlates slightly better than ASR_{EIS} with the absolute average discharge cell potential (V_{dc}) during cycling (averaged over 19 cycles at 25 mA cm⁻²) (Fig. 10b). However, discrepancies here also highlight that flow cell performance, as assessed by polarisation or impedance, does not necessarily predict all aspects of cell CD cycling behaviour, even under identical configurations. For instance, EU vs. ASR_{pol} are effectively uncorrelated (Fig. 10c). While a highly resistive cell may reach the cut-off voltage prematurely, cut-off is often instead governed by a sharp voltage spike due to depletion of the active species. Thus, factors that have little influence on polarisation behaviour at 50% SoC (*e.g.* tank geometry, mixing, or other design-specific parameters) may exert a greater effect on EU. This illustrates that EU is, to some extent, independent of cell polarisation characteristics, being influenced instead by broader system-level and operational factors.

Scatterplots between pairs of EU, CE, average charge-discharge voltage, R_{Ω} , R_{CT} , R_{MT} , ASR_{pol} , ASR_{EIS} , iR -corrected ASRs, and decay rate suggest R_{Ω} largely influences cell efficiency techniques, and R_{CT} and R_{MT} have an outsized influence on a subset of the data (SI Fig. S10 and S11). Most pairs of these variables weakly or do not correlate, potentially evincing multiple sources of variability. Generally, these relationships highlight how single techniques (polarisation, impedance), if not carefully considered, are imperfect proxies for CD cycling performance, especially when performing interlaboratory comparisons. Such discrepancies might arise from inadequate assumptions about connections between data from different techniques, due to time-dependent system properties, and/or due to different phenomena affecting performance of individual cells.

Hypothesised and tested influencing factors

In Table 1, we present statistical measures of the metrics explored in this work (focusing on the datasets collected at

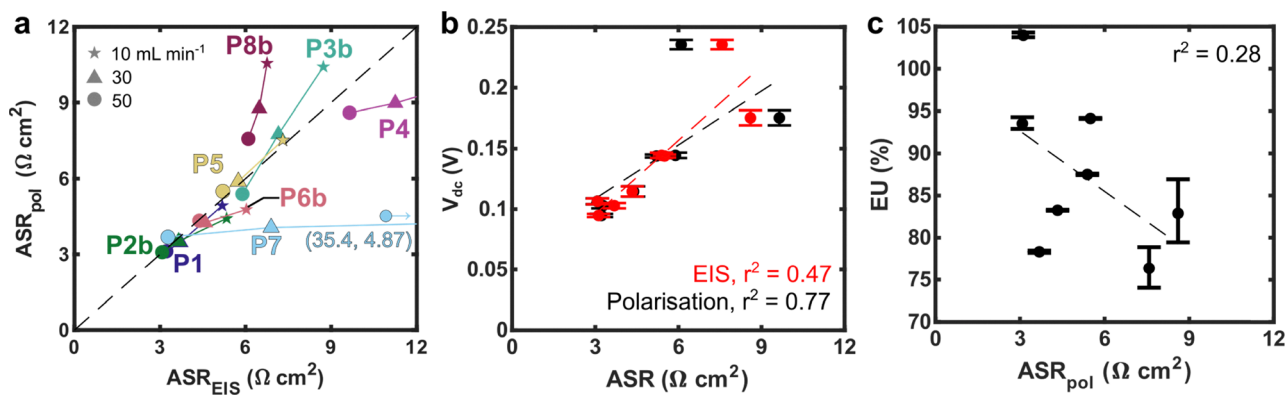


Fig. 10 Correlations between datasets. (a) ASR obtained from slope of polarisation (ASR_{pol}) curve (up to 25 mA cm⁻²) vs. that obtained from impedance (ASR_{EIS}). A dashed line indicates where $ASR_{pol} = ASR_{EIS}$ obtained at 10, 30, and 50 mL min⁻¹. (b) Average discharge cell potential during cycling (at 25 mA cm⁻²) vs. polarisation- or EIS-derived ASR at 50 mL min⁻¹. (c) Electrolyte utilisation (EU) vs. polarisation-derived ASR at 50 mL min⁻¹. In (b) and (c), the error bars represent the standard deviation in the average discharge potential or EU per cycle ($N = 19$ cycles, 1 cell per participant).



Table 1 Statistics for excluding group (excluding P3b, P4, P7, and P8b, $N = 4$ participants). ASRs are in units of $\Omega \text{ cm}^2$ and are the values collected at 50 mL min^{-1} . Parameter meanings: mean (μ), standard deviation (σ), inter-quartile range (IQR), 25th quartile (Q25), 75th quartile (Q75). Complete tables representing these datasets (and all data) are included in the SI Table S6. *Implies excluding P1, P4, and P8b ($N = 5$ participants). P7, P3b are not excluded because the reason for their exclusion in polarisation/impedance (SoC drift) is irrelevant here. P1 is excluded due to the excess electrolyte volume

	μ	σ	IQR	Q75	Q25	Median
ASR _{pol}	4.01	1.15	1.81	4.92	3.10	3.73
ASR _{EIS}	3.97	1.01	1.65	4.79	3.15	3.80
R_{Ω} ($\Omega \text{ cm}^2$)	2.22	0.84	1.27	2.86	1.59	2.00
R_{CT} ($\Omega \text{ cm}^2$)	0.47	0.34	0.47	0.70	0.23	0.48
R_{MT} ($\Omega \text{ cm}^2$)	1.28	0.31	0.52	1.54	1.02	1.27
ASR _{pol,iR-corr}	1.79	0.33	0.55	2.06	1.51	1.76
ASR _{EIS,iR-corr}	1.75	0.23	0.38	1.94	1.56	1.72
*EU (%)	87.37	6.77	11.70	93.73	82.03	87.52
*CE (%)	99.87	0.11	0.17	99.96	99.80	99.88
* V_{ch} (V)	0.123	0.020	0.039	0.144	0.106	0.115
* V_{dch} (V)	0.124	0.020	0.037	0.144	0.107	0.116
*Decay ($\% \text{ day}^{-1}$)	-0.32	0.43	0.58	-0.07	-0.65	-0.16

50 mL min^{-1} and excluding outliers where relevant). In this group, ASR_{pol} ($4.01 \pm 1.15 \Omega \text{ cm}^2$) is nearly equivalent to ASR_{EIS} ($3.97 \pm 1.01 \Omega \text{ cm}^2$) in both mean and variance metrics. Most of this error can be attributed to ASR _{Ω} ($2.22 \pm 0.84 \Omega \text{ cm}^2$) whereas the remaining error estimated with *iR*-corrected polarisation (ASR_{pol,iR-corr}, $1.79 \pm 0.33 \Omega \text{ cm}^2$) and the combined portion, ASR_{MT} + ASR_{CT} (ASR_{EIS,iR-corr}, $1.75 \pm 0.23 \Omega \text{ cm}^2$), are comparable in magnitude, but of lesser variance. That the variance of the individual ASR_{MT} ($0.47 \pm 0.34 \Omega \text{ cm}^2$) and ASR_{CT} ($1.28 \pm 0.31 \Omega \text{ cm}^2$) are comparable or even higher than ASR_{pol,iR-corr} or ASR_{EIS,iR-corr} suggests that some part of the error derives from the fitting or interpretation of the impedance data with the specified equivalent circuit elements. Regarding CD cycling, the largest errors are observed for EU ($87.37 \pm 6.77\%$) and the variance in the charge (V_{ch})/discharge (V_{dch}) potential ($\pm 20 \text{ mV}$) is similar to that observed in polarisation (Fig. 5, 27 mV at 25 mA cm^{-2}). Finally, as discussed previously, decay rates span a large range (-0.066 to $-1.1\% \text{ day}^{-1}$, excluding P1, P4, and P8b).

Broad differences in choices or environmental factors of the system set-up (Table S3), electrochemical protocol (Table S4), and during cell operation (Table S5), and the relatively low number of experimental data ($N = 8$ participants), obfuscate attribution of these factors to specific performance outcomes. For instance, no correlations were observed between the average ambient temperature and any performance metrics (for any subsets of data excluding or including outliers), reflecting uncertainties in the measurement and that other effects dominate performance. An insufficient number of participants could be grouped together to distinguish performance between electrical connector type (crocodile clip – 5, Pomona connector – 1, brass screw – 1, excluding P4 here due to reported current collector fouling) or for different pump calibration procedures. Given the relative insensitivity to flow rate between $30\text{--}50 \text{ mL min}^{-1}$ over 0 to 25 mA cm^{-2} (Fig. S8), and the greater apparent

influence of ohmic losses across most cells, we believe pump calibration does not explain performance differences across participants (although there is the possibility of unknown incorrect flow rates). We also attempted to group participants into those that conditioned their cell using electrochemistry – 2, pumping electrolyte (at OCV for variable durations) – 4, or none – 1 (again, excluding P4 from this analysis). However, these treatments were distinct in protocol and “break-in” effects might occur during the earlier flow cell measurements. Electrode cutting was also difficult to quantify and cable length did not correlate highly with any performance metrics. While the short lengths of cabling should introduce minimal resistance (typical 0.75 mm^2 or 4 mm^2 copper wires contribute resistances of *ca.* 25 to $5 \text{ m}\Omega \text{ m}^{-1}$, respectively), these translate to ASRs comparable to those of the flow cell (*e.g.*, a 2.5 m length, 0.75 mm^2 cross-section wire approximately contributes $62.5 \text{ m}\Omega$ resistance, which would correspond to $1 \Omega \text{ cm}^2$ ASR). Larger flow cells are thus generally expected to suffer more from stray resistances. Similarly, contact resistances at the leads are expected to be a larger portion of the resistance in low-resistance large-area cells. Because ohmic losses (estimated with impedance) and EU had the largest noticeable differences, and because there are more-detailed treatments of ferri-/ferro-cyanide decay processes,⁴⁰ we elected to test hypotheses on these two metrics.

Given the aforementioned sensitivity of the measured ASR to contact effects and noting that participants employed the same flow cell architecture and component set, we hypothesised that a non-negligible fraction of the variance in ohmic losses arose from differences in electrical connection practice at the cell terminals. To assess this contribution directly, we measured the resistance of a dry cell, defined here as a fully assembled cell without a membrane and without electrolyte filling, thereby isolating electronic and contact resistances from ionic contributions. Six wiring and connection configurations were evaluated to reflect approaches adopted by participants, as well as other reasonable connection methods, which are depicted in Fig. 11a. In configuration (i), the potentiostat leads were shorted to provide a baseline for the measurement system. Configurations (ii) to (v) employed manufacturer-supplied potentiostat cables (Biologic) with variations in how current-carrying and voltage-sensing leads were attached at the current collectors. Configuration (vi) employed proprietary cable extensions. In all instances where a four-point connection was employed, the voltage-sensing cable was attached to the stem of the current collector (Fig. S2a) using a crocodile clip.

Across these configurations, two-point connections, in which the voltage-sensing and current-carrying leads were joined prior to attachment at the cell, yielded markedly higher measured resistance (particularly configuration (iv), which connected to a Pomona connector on the cell) than four-point connections. This is consistent with the inclusion of additional contact and lead resistances in the voltage measurement. By comparing the shorted lead baseline with the four-point connected dry cell, we estimate a dry cell resistance of *ca.* $0.1 \Omega \text{ cm}^2$ (Fig. 11b). Configuration (iv) introduced an





Fig. 11 Configurations and derived metrics from testing hypothesised influencing factors. (a) Schematic showing cell connection configurations evaluated: (i) potentiostat cables directly shorted; (ii) two-point connection attached to current collectors *via* crocodile clips; (iii) four-point connection attached to current collectors *via* crocodile clips; (iv) two-point connection attached to current collectors *via* crocodile clips/Pomona connectors; (v) four-point connection attached current collectors *via* crocodile clips (voltage) and *via* crocodile clips/Pomona connectors (current); and (vi) four-point connection attached to current collectors *via* crocodile clips (voltage) and *via* proprietary cable extensions using Pomona connectors (current). (b) Bar chart showing ASRs of the different cell connection configurations. (c) Schematic illustrating tank set-up and configuration variables. (d) Schematics delineating tubing/mixing configurations examined at coordinating institutions. (e) Grouped scatter plot showing three-cycle EU average values and their standard deviations (error bars in black) when cells were CD cycled with different tubing/mixing configurations.

additional $0.90 \Omega \text{ cm}^2$, which corresponds to *ca.* 24% of the median ASR (ASR_{pol} or ASR_{EIS}) reported in Table 1. This magnitude is therefore shown to be sufficient to account for an appreciable proportion of the interlaboratory spread in reported ohmic losses. Consistent with this interpretation, participants employing a four-point connection reported a mean R_{Ω} of $1.85 \pm 0.45 \Omega \text{ cm}^2$ (P1, P2 and P6b, at 50 mL min^{-1}) as shown in Fig. 6d) whereas the remaining participants, who employed a two-point connection (P3b, P4, P5, P7 and P8), reported a higher mean of $3.09 \pm 0.87 \Omega \text{ cm}^2$. Notably, in the dry cell measurements reported here, introducing cable extensions and intermediate connectors (configuration (vi)) did not measurably increase the ASR when a four-point connection was maintained, relative to the comparable configuration without extensions. This result is specific to the configurations tested and should not be interpreted as a general statement about all extension hardware. Rather, it indicates that four-point connection can mitigate the influence of additional series resistances external to the cell, rendering the measured R_{Ω} less sensitive to cable extensions than under two-point connection. Additional resistance remaining even under 4-point connections might be associated with variance in internal contact resistances, and variance in membrane resistance due to storage, handling, history prior to measurement, and material heterogeneity.

Elsewhere in the study, EU exhibited substantially greater interlaboratory variability than other CD cycling-derived metrics (Fig. 8), motivating additional experiments to evaluate

whether reservoir level mixing effects could contribute. We hypothesised that, depending on the reservoir geometry and mixing regime, fluid bypass between the inlet and outlet could reduce the effective exchange of electrolyte between the cell and the bulk reservoir. In this scenario, electrolyte local to the tubing ends may not mix effectively with the bulk reservoir, thereby diminishing the charge and discharge capacities accessed during CD cycling.

To examine this hypothesis, additional experiments were performed at two coordinating institutions. At Institution 1, the influence of stirring was evaluated using three regimes: (i) high-speed stirring at *ca.* 1000 rpm; (ii) no stirring; and (iii) low-speed stirring at *ca.* 150 rpm. At Institution 2, testing was performed with the outlet tube above the electrolyte or submerged in it, each with and without stirring, giving the following four configurations: (iv) outlet above, with stirring; (v) outlet above, without stirring; (vi) outlet submerged, without stirring; and (vii) outlet submerged, with stirring. All seven configurations are illustrated in Fig. 11c and d. Experiments were carried out in 100 mL borosilicate glass media bottles, for three cycles, and each experiment was repeated three times. The three-cycle average EU value *per* repeat of each tubing/mixing configuration are shown in Fig. 11e.

EU was comparable across the two coordinating institutions. Institution 1 reported an average EU of $86.9 \pm 0.68\%$ for configuration (iii), whereas Institution 2 reported an average EU of $90.8 \pm 2.34\%$ for the corresponding condition (configuration (iv)). While the mean EU was slightly lower for



Institution 1, the repeat measurements were more tightly grouped. It is notable that the researcher at Institution 1 assembled the cells for all three repeats at the same time and operated them simultaneously. The researcher at Institution 2 carried out the cell repeats one after the other, with both assembly and operation occurring across multiple days. For most configurations, EU remained in the range of *ca.* 86 to 93%. However, when stirring was off and the tubing ends were close together (both submerged in electrolyte as shown in configuration (vi)), EU decreased markedly, to an average of $30.4 \pm 8.70\%$. By contrast, switching on stirring increased EU to values comparable to those obtained with separated tubing, even when the tubing ends remained in proximity. These results indicate that inlet and outlet placement and reservoir mixing conditions can significantly influence EU, with the magnitude of the effect expected to depend on reservoir geometry and operating conditions. Related utilisation and reservoir mixing effects have been discussed in greater detail for vanadium flow battery systems.^{60,61}

The effects studied in this section are of common consideration across flow battery chemistries. For example, electrical lead connections should be optimised to impart the minimum resistance, regardless of electrolyte chemistry, and ensuring the electrolyte is homogeneously mixed within reservoirs should serve to minimise error in EU measurements. Other effects, which we did not explore in more depth, are expected to be sensitive to a host of factors. From the perspective of certain measurable parameters, for instance, R_{MT} and R_{CT} are expected to be sensitive to electrode choice, its pre-treatment, its functionalisation (*e.g.*, decoration with catalyst), and the redox chemistry of interest. Moreover, capacity decay is also likely governed largely by a host of chemistry-specific considerations. Therefore, the sensitivities of each system to considerations such as flow-rate calibration and electrode cutting procedure are likely to require further study.

Recommendations for experiments and reporting

The following suggestions are based on observations reported in this manuscript and are not intended to be comprehensive. Other redox chemistries and operating conditions may elicit different sensitivities to these (and other) factors.

(1) Develop polarisation protocols which do their best to: (a) start from a consistent SoC for each data point (*i.e.*, by Coulomb counting back to original SoC), (b) collect data beyond early-time transients (*e.g.*, due to boundary-layer development of the electrolyte, although other phenomena may contribute to such transients), and (c) minimise the influence of SoC drift by balancing system reservoir size with step duration (eqn (1) and (2)). Report the polarisation protocol thoroughly (SoC, step durations, sequences, SoC compensation) and data processing details (*e.g.*, which portion of the data is averaged to produce the polarisation plot). If the system exhibits unexpected transient behaviour, it may be worth reporting those traces. Measure and report OCV at the beginning and end of each polarisation set to gauge SoC drift.

Reporting *iR*-corrected polarisation may also be worthwhile, particularly for explorations into electrode performance.

(2) Use a 4-point probe configuration to minimise the influence of connection contact resistances and cables on ohmic losses. Larger-area cells will be more sensitive to “stray” resistances due to cabling and connections. Measuring resistances across the cell in the absence of electrolyte and membrane (a dry cell) may be worthwhile for setting expectations of resistances across connections. Note, however, that 4-point connections may also be more sensitive to high-frequency inductive artefacts, depending on the connection configuration.

(3) Validate operational parameters and material functions. For flow rates, simple checks, such as volumetric or gravimetric measurements over a fixed duration can ensure a specified flow rate is achieved. In some pumping configurations, it may be important to measure the flow rate with cell and other hardware connected inline. Visual inspection of the flow cell components after experiments can help in identifying leaks, damaged or degraded components (*e.g.*, compromised current collectors), and electrolyte decomposition (*e.g.*, colour changes to electrolyte at same SoC).

(4) Evacuate air pockets from the system. Agitating and inverting the flow cell whilst flowing electrolyte can help evacuate bubbles that may limit electrode and/or membrane wetting.

(5) For electrolyte replacement, evacuate the contents entirely from the cell as it may influence parameters of interest. Exchanging electrolyte by only pumping out fluid, especially when the flow cell volume is relatively large relative to that of the reservoir, may leave an appreciable electrolyte volume in the cell that has an outsized effect on parameters such as EU. Flush with copious DI water and then remove bulk DI water (*e.g.*, with inert gas, taking precautions to not splash the operator).

(6) Sparge and blanket with humidified inert gas. Dry gas may cause solvent evaporation, thus modifying electrolyte properties. First, use a low gas flow rate to sparge (remove dissolved air) the electrolyte and then blanket the headspace to minimise agitating the electrolyte. Aggressive sparging may deposit droplets on the internal reservoir walls leading to capacity loss, particularly for low-volume systems.

(7) Holistically report ancillary equipment and operation details for sparging gas (if using, provide purity, pre-humidification, flow rate), reservoir stirring (ideally stir to minimise low EU from inadequate mixing), materials which contact the electrolyte (to assist others in material selection or to identify incompatibilities), and pump (model, physical limitations such as flow rate, maximum pressure).

(8) Evaluate “break-in” and follow a repeatable “break-in” protocol. The system may drift in performance with time upon start-up due to a variety of phenomena. As such, repeated tests (*e.g.*, impedance, polarisation) whilst pumping electrolyte can elucidate this and ensure a similar starting point for experiments. Report “break-in” protocols, if they are used, to help raise awareness of such system behaviours.

(9) Consider purpose-designed systems to measure properties with single-electrolyte setups. Set-ups and techniques exist



to specifically measure properties of interest with minimal influencing factors (*e.g.*, volumetrically imbalanced symmetric flow cells employing constant current followed by constant voltage techniques for capacity decay of certain active species). Employ single reservoir set-ups when performing EIS or polarisation studies of a redox electrolyte to minimise SoC drift during measurements.^{41,46,62}

(10) Repeat experiments when possible. This serves to gather critical information about repeatability which is currently lacking in the literature. Report averaged metrics with standard deviations across a defined number of repeats.

(11) Unambiguously specify the electrolyte by stating the individual component concentrations and the volumes in each reservoir. For instance, “a 100 mM ferri-/ferro-cyanide solution at 50% SoC” might be construed as either 100 mM or 50 mM of each of ferri- and ferro-cyanide. “100 mM ferricyanide and 100 mM ferrocyanide” is clearer. “100 mL of electrolyte” should read “100 mL of electrolyte in each reservoir”. Ultimately, the electrolyte should be able to be prepared to the same specification unambiguously, which for some chemistries may require consideration of purity specification.

Round robin design recommendations

Our review of this data, a follow-up workshop with participants, and reviewer comments helped us identify recommendations for future round-robin studies. Here we provide these within a stepwise procedure for performing a round-robin study of flow cells or batteries.

(1) Conceptualise the round-robin study and recruit participants. At a high level you need a chemistry, flow cell device, participants, question(s) of interest, and communication. A conference both inspired our study and provided participants. A subset from the conference proceeded to develop the study. Since, we have found that crowdsourcing is useful to project scoping. In our study, we limited information sharing to minimise collective influencing of test results. However, this led to hesitation in engagement even after data were collected, limiting discussions and connections between researchers. Post-experiment discussions with participants suggested that clearer communication could have also clarified some request details. As such, we believe that round robins benefit from open communication and, as collaborative exercises, can be used for community building.

(2) Prepare a preliminary experimental request. Identify the variables of interest and what should be explicitly controlled. Maintain consistency in configuration throughout experiments where possible (*e.g.*, do not change lead configuration) to minimise comparative complications between different experiments. Determine how variables can be quantified (*vs.* those that might be considered categorical), and whether it is practical for participants to measure quantitative variables. Some variables are best measured during operation (*e.g.*, temperature). Include guidelines for material pre-treatment and storage. The request should be concise, but detailed. It should also be structured to ensure thorough reporting of experimental parameters, sequence, and timing. Develop surveys to collect

post-experiment data and experimental practices. In the experimental request, include repeats to distinguish between repeatability and replicability errors and to capture changes between sequential experiments. Incorporate criteria for acceptable data or experimental results (*e.g.*, require re-collection of data for leaky or compromised cells, validation of flow rate). Such “quality checks” improve the likelihood that the effects of the intended explored parameters are studied. Incorporate as much control as possible if the intent is to explore specific effects. In our round robin, for instance, minimising exposure to light would have lessened concerns over light-induced capacity-loss to focus on variability in CD cycling performance due to other factors. Additionally, limiting the data collection to the same current density range would have resulted in a simpler dataset to process (no optional additional data). Prepare a system to anonymise data. Be clear about how data might be shared between participants and broader communities.

(3) Survey for capabilities and typical practices. In our surveys, there were opportunities to collect additional information (*e.g.*, electrical leads configuration, clarity on the timing of experiments, inert gas identity and purity). Better-designed surveys with fewer free-response answers could have likely accelerated their processing. Additionally, greater diligence in processing the surveys may have prevented oversight of requesting unfeasible conditions (*e.g.*, 1 mL min⁻¹ flow rate).

Ensure that the experimental request is not limited by equipment. Use the preliminary experimental request to guide questions (*e.g.*, access to specific techniques, flow rate capabilities, current and voltage range of potentiostats, specific flow cell equipment, reservoir volumes). Be quantitative where possible. More information is better, although this will need to be balanced against ease-of-processing and willingness of participants to fill out lengthy surveys. Processing time will scale both with the information requested and the number of participants. A well-motivated, organised, and concise survey will simplify analysis. Structure the forms to be multiple-choice (or numerically enforced) where possible to constrain answers. Break down questions into multiple ones to minimise free-response answers that otherwise require substantial digestion. For instance, “What flow cell do you use?” is better broken down into multiple questions that target the specific flow cell components. Provide examples for each free-response question to encourage consistent answers.

(4) Have the coordinating team test the experimental protocol to gather some expectations for performance. Challenges in experimentation are easier to correct before sharing the experimental request. This is a good way to additionally develop the quality checks and to observe whether processes like “break-in” might influence performance. Having multiple participants perform this step can surface issues not captured by a single researcher (*e.g.*, due to equipment or experience differences).

(5) Refine the experimental request. Use the survey information to update parameter choices and ensure the timescales of the experimental request are reasonable. Share the refined request with participants to gather feedback before requesting data. Revisit the surveys to capture post-experiment data and



the experimental practices themselves. Because different potentiostats and other equipment may be used for data collection, provide templates for data input to ease analysis. Communicate if there are multiple rounds of data collection and if there is an intended “refinement” period to catch experimental or protocol issues. Try to clarify expectations and timelines as much as possible.

(6) Launch the experimental request. Encourage and be prepared to handle feedback (*e.g.*, fixing unclear instructions). Plan for timeline extensions to handle contingencies and researcher needs.

(7) Collect and process data centrally and consistently. We received data in different formats that were processed in three different software tools (Excel, MATLAB, Python) by three different authors. We could have improved the efficiency of data processing by standardising submission data formats (*e.g.*, .CSV with same time formats and variable names), defining and harmonising data processing workflows earlier in the study, and creating sharable repositories for the data and processing code. A central repository can facilitate collaborative data processing. Ensure that data are made anonymous if shared. Develop scripts which process data consistently. Also, aligning early on which correlations to explore can accelerate processing.

(8) Develop and test hypotheses for performance-influencing factors. Involve participants to maximise information- and hypothesis-sharing to explain results. Determine whether clear correlations arise across performance and operation choices. Design experiments to test these correlations. In our study, a large set of potentially-influencing factors made attribution between factors and performance difficult. Follow-up testing in this case involved fixing all possible variables and testing variables that we hypothesised to be consequential. Ideally, with relatively short turnaround times, participants can help assess the influence of individual parameters.

Summary

In this round-robin study, we measured replicability in flow cell performance across eight participating research groups who received the same experimental request to conduct three common electrochemical techniques using the same model ferri-/ferro-cyanide-based electrolyte, cell architecture, and component material sets. The experimental request, intended to have a similar level of detail as a “Methods” section in flow battery literature, left reasonable room for researchers to interpret how they set up and tested their flow cells. Through surveys, we documented some of their individual decisions, showing differences in balance-of-system set-up, experimental procedures, and cell operation. Of note, we found that polarisation protocols vary significantly between participants, whereas ultimately these disparate approaches are transformed into the same representation (polarisation curves). Subsequently, we discussed the variable performance in polarisation, impedance, and CD cycling. Seemingly, similar polarisation data (~ 15 mV standard deviation in

potential at 60 mA cm^{-2}) can be obtained through *iR*-correction for the cells with comparable polarisation protocols. Lesser agreement in the absence of *iR*-correction suggests opportunities in addressing variation in ohmic losses. The data gathered using EIS revealed substantial differences in the Nyquist plots and consequently in their fits to the same equivalent circuit model. Parameters extracted from these fits generally followed some expected trends (*i.e.*, kinetic and ohmic resistances were largely independent of flow rate, whereas mass transport resistance depended on flow rate); however, the absolute parameter values varied significantly between participants. When analysing CD cycling data, significant differences were observed between participants across most calculated metrics, including EU, average charge-discharge voltages, and capacity decay rate. The extent of variability differed significantly among these parameters, with electrolyte utilisation (EU) exhibiting the largest average standard deviation, approaching 10%. Further, as anticipated, metrics suggesting better-performing behaviour in polarisation positively correlated with metrics derived from impedance and CD cycling. However, these relationships were somewhat limited in their predictive power for the conditions studied. Due to the breadth of differences in operational and system choices, in addition to variation in environmental factors (*e.g.*, reservoir sealing, electrolyte exposure to light) or history-dependent behaviours (*e.g.*, experiment ordering and protocol discrepancies), we cannot unequivocally attribute the extent of performance differences to specific experimentalist choices. Nonetheless, we hypothesised a few parameters to be of consequence to the observed performance and through additional experiments found that ohmic losses were non-negligibly influenced by electrical connection configuration (measured variability corresponded to *ca.* 24% of the median total cell ASR) and that EU could be dramatically reduced by poor electrolyte mixing resultant from fluid bypass in reservoirs. This study has also highlighted other operational and system choices, *e.g.* flow rate verification, “break-in” procedures (including cell priming), and electrochemical technique sequencing, that spotlight opportunities for further investigation and serves as a first step towards improving methodological transparency and developing flow cell testing standards. The design, outcomes, and lessons learnt from this interlaboratory study provide the blueprints for future round-robin studies, whilst establishing deeper interlaboratory connections that encourage community growth and discussions on factors of flow cell performance beyond those communicated through the typical channels. Given the level of variability identified, with a simple chemistry and the same cell, provided to researchers in established and well-resourced laboratories with flow battery experience, it seems clear that the field in general would benefit from further investigations of this kind.

Author contributions

Hugh O'Connor, Alexander H. Quinn, Fikile R. Brushett, and Josh J. Bailey conceived, organised, and administrated the project. Hugh O'Connor manufactured and disseminated the flow cell and materials kits. Hugh O'Connor, Alexander



H. Quinn, Edward Saunders, Aodhán Dugan, Thomas Goodwin, Nadia Farag, Greta Thompson, Ameya Bondre, Marina Tabuyo-Martinez, Hannah M. Burnett, T. Y. George, Jordan D. Sosa, and Carlos Mingoos supplied survey information and collected the flow cell data. Hugh O'Connor, Alexander H. Quinn, and Josh J. Bailey analysed survey information and drafted the initial manuscript. Josh J. Bailey anonymised the data before it was analysed. Hugh O'Connor processed and analysed the CD cycling data. Edward Saunders analysed the impedance data. Alexander H. Quinn processed and analysed the polarisation data and processed the impedance data. Fikile R. Brushett and Josh J. Bailey reviewed and edited all manuscript drafts. Peter Nockemann, Fikile R. Brushett, Clare P. Grey, Dominic Wright, Michaël De Volder, Antoni Forner-Cuenca, Robert A. W. Dryfe, Michael J. Aziz, Ana B. Jorge Sobrido, and Josh J. Bailey reviewed the final drafts and provided supervision. Josh J. Bailey provided project funds.

Conflicts of interest

There are no conflicts to declare.

Data availability

Data are provided in the supplementary information (SI). SF1 is the pre-experiment survey, SF2 is the experimental request, SF3 is the post-experiment practices survey, SF4 is the post-experiment data analysis and reporting survey, SF5 is a .zip containing files relating to 3D-printing, SF6 is a .zip containing files relating to computer-aided design, SF7 is the cell assembly guide. The electrochemical data are found in the SF8 "Data and code.zip" file. See DOI: <https://doi.org/10.1039/d5ee07103h>.

Acknowledgements

The authors gratefully acknowledge the Queen's University Belfast Agility Fund+ scheme for funding research activities. All participants greatly acknowledge the RSC Researcher Collaborations Grant (C24-8470737976) for support. FRB and AHQ gratefully acknowledge support from The Royal Society International Exchanges Grant (IES\R3\213001). AHQ gratefully acknowledges the National Science Foundation Graduate Research Fellowship Program under Grant Number 1745302. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF. AHQ also acknowledges the Alfred P. Sloan Foundation's Minority PhD (MPHD) Program. ABJS acknowledges the UK Research and Innovation for her Future Leaders Fellowship no. MR/T041412/1. Fig. 1 was created using BioRender, incorporating illustrations generated with ChatGPT-4 (OpenAI) based on original photographs and descriptions from the study to produce stylised depictions of experimental activities.

References

- 1 L. Aspitarte and C. R. Woodside, *Cell Rep. Sustainability*, 2024, **1**, 100007.
- 2 V. Sprenkle, B. Li, L. Zhang, L. A. Robertson, Z. Li and P. Balducci, *Technology Strategy Assessment: Flow Batteries, DOE/OE-0033*, U.S. Department of Energy, Office of Electricity, 2023.
- 3 A. Lucas and S. Chondrogiannis, *Int. J. Electr. Power Energy Syst.*, 2016, **80**, 26–36.
- 4 M. Mansha, A. Ayub, I. A. Khan, S. Ali, A. S. Alzahrani, M. Khan, M. Arshad, A. Rauf and S. Akram Khan, *Chem. Rec.*, 2024, **24**, e202300284.
- 5 M. Mansha, A. Anam, S. Akram Khan, A. Saeed Alzahrani, M. Khan, A. Ahmad, M. Arshad and S. Ali, *Chem. Rec.*, 2024, **24**, e202300233.
- 6 R. Khezri, P. Tangthum, A. A. Mohamad and S. Kheawhom, in *Electrochemical Energy Storage Technologies Beyond LI-ION Batteries*, ed. G. He, Elsevier, 2025, pp. 461–477.
- 7 S. Zhang, S. Gao, Y. Zhang, Y. Song, I. R. Gentle, L. Wang and B. Luo, *Energy Storage Mater.*, 2025, **75**, 104004.
- 8 G. Tian, A. J. Sobrido, L. M. M. Herrera, S. Doszczeczko and M. W. Thielke, in *Electrochemical Energy Storage Technologies Beyond LI-ION Batteries*, ed. G. He, Elsevier, 2025, pp. 427–459.
- 9 L. Zhang, R. Feng, W. Wang and G. Yu, *Nat. Rev. Chem.*, 2022, **6**, 524–543.
- 10 M. Park, J. Ryu, W. Wang and J. Cho, *Nat. Rev. Mater.*, 2016, **2**, 1–18.
- 11 C. Zhang, Z. Yuan and X. Li, *ACS Energy Lett.*, 2024, **9**, 3456–3473.
- 12 D. S. Aaron, Q. Liu, Z. Tang, G. M. Grim, A. B. Papandrew, A. Turhan, T. A. Zawodzinski and M. M. Mench, *J. Power Sources*, 2012, **206**, 450–453.
- 13 Q. Xu, T. S. Zhao and C. Zhang, *Electrochim. Acta*, 2014, **142**, 61–67.
- 14 V. Muñoz-Perales, P. Á. García-Salaberri, A. Mularczyk, S. Enrique Ibáñez, M. Vera and A. Forner-Cuenca, *J. Power Sources*, 2023, **586**, 233420.
- 15 X. Ke, J. M. Prah, J. I. D. Alexander, J. S. Wainright, T. A. Zawodzinski and R. F. Savinell, *Chem. Soc. Rev.*, 2018, **47**, 8721–8743.
- 16 H. O'Connor, A. H. Quinn, F. R. Brushett, O. Istrate, S. Glover, J. J. Bailey and P. Nockemann, *Discover Electrochem.*, 2025, **2**, 23.
- 17 G. Smith and E. J. F. Dickinson, *Nat. Commun.*, 2022, **13**, 6832.
- 18 M. Baker, *Nature*, 2016, **533**, 452–454.
- 19 E. National Academies of Sciences, *Reproducibility and Replicability in Science*, 2019.
- 20 S. L. McArthur, *Biointerphases*, 2019, **14**, 020201.
- 21 A. K. Stephan, *Joule*, 2021, **5**, 1–2.
- 22 I. Rios Amador, R. T. Hannagan, D. H. Marin, J. T. Perryman, C. Rémy, M. A. Hubert, G. A. Lindquist, L. Chen, M. B. Stevens, S. W. Boettcher, A. C. Nielander and T. F. Jaramillo, *STAR Protoc.*, 2023, **4**, 102606.



- 23 J. Li, C. Arbizzani, S. Kjelstrup, J. Xiao, Y. Xia, Y. Yu, Y. Yang, I. Belharouak, T. Zawodzinski, S.-T. Myung, R. Raccichini and S. Passerini, *J. Power Sources*, 2020, **452**, 227824.
- 24 Y.-K. Sun, *ACS Energy Lett.*, 2021, **6**, 2187–2189.
- 25 K. Ehelebe, N. Schmitt, G. Sievers, A. W. Jensen, A. Hrnjić, P. Collantes Jiménez, P. Kaiser, M. Geuß, Y.-P. Ku, P. Jovanović, K. J. J. Mayrhofer, B. Etzold, N. Hodnik, M. Escudero-Escribano, M. Arenz and S. Cherevko, *ACS Energy Lett.*, 2022, 816–826.
- 26 G. Bender, M. Carmo, T. Smolinka, A. Gago, N. Danilovic, M. Mueller, F. Ganci, A. Fallisch, P. Lettenmeier, K. A. Friedrich, K. Ayers, B. Pivovar, J. Mergel and D. Stolten, *Int. J. Hydrogen Energy*, 2019, **44**, 9174–9187.
- 27 S. Ritter, R.-W. Bosch, F. Huet, K. Ngo, R. A. Cottis, M. Bakalli, M. Curioni, M. Herbst, A. Heyn and J. Macak, *Corros. Eng., Sci. Technol.*, 2021, **56**, 254–268.
- 28 R.-W. Bosch, R. A. Cottis, K. Csecs, T. Dorsch, L. Dunbar, A. Heyn, F. Huet, O. Hyökyvirta, Z. Kerner, A. Kobzova, J. Macak, R. Novotny, J. Öijerholm, J. Piippo, R. Richner, S. Ritter, J. M. Sánchez-Amaya, A. Somogyi, S. Väisänen and W. Zhang, *Electrochim. Acta*, 2014, **120**, 379–389.
- 29 J. W. Gittins, Y. Chen, S. Arnold, V. Augustyn, A. Balducci, T. Brousse, E. Frackowiak, P. Gómez-Romero, A. Kanwade, L. Köps, P. K. Jha, D. Lyu, M. Meo, D. Pandey, L. Pang, V. Presser, M. Rapisarda, D. Rueda-García, S. Saeed, P. M. Shirage, A. Ślesiński, F. Soavi, J. Thomas, M.-M. Titirici, H. Wang, Z. Xu, A. Yu, M. Zhang and A. C. Forse, *J. Power Sources*, 2023, **585**, 233637.
- 30 J. W. M. Osterrieth, J. Rampersad, D. Madden, N. Rampal, L. Skoric, B. Connolly, M. D. Allendorf, V. Stavila, J. L. Snider, R. Ameloot, J. Marreiros, C. Ania, D. Azevedo, E. Vilarrasa-García, B. F. Santos, X.-H. Bu, Z. Chang, H. Bunzen, N. R. Champness, S. L. Griffin, B. Chen, R.-B. Lin, B. Coasne, S. Cohen, J. C. Moreton, Y. J. Colón, L. Chen, R. Clowes, F.-X. Coudert, Y. Cui, B. Hou, D. M. D'Alessandro, P. W. Doheny, M. Dincă, C. Sun, C. Doonan, M. T. Huxley, J. D. Evans, P. Falcaro, R. Ricco, O. Farha, K. B. Idrees, T. Islamoglu, P. Feng, H. Yang, R. S. Forgan, D. Bara, S. Furukawa, E. Sanchez, J. Gascon, S. Telalović, S. K. Ghosh, S. Mukherjee, M. R. Hill, M. M. Sadiq, P. Horcajada, P. Salcedo-Abraira, K. Kaneko, R. Kukobat, J. Kenvin, S. Keskin, S. Kitagawa, K. Otake, R. P. Lively, S. J. A. DeWitt, P. Llewellyn, B. V. Lotsch, S. T. Emmerling, A. M. Pütz, C. Martí-Gastaldo, N. M. Padial, J. García-Martínez, N. Linares, D. Maspoch, J. A. Suárez del Pino, P. Moghadam, R. Oktavian, R. E. Morris, P. S. Wheatley, J. Navarro, C. Petit, D. Danaci, M. J. Rosseinsky, A. P. Katsoulidis, M. Schröder, X. Han, S. Yang, C. Serre, G. Mouchaham, D. S. Sholl, R. Thyagarajan, D. Siderius, R. Q. Snurr, R. B. Goncalves, S. Telfer, S. J. Lee, V. P. Ting, J. L. Rowlandson, T. Uemura, T. Iiyuka, M. A. van der Veen, D. Rega, V. Van Speybroeck, S. M. J. Rogge, A. Lemaire, K. S. Walton, L. W. Bingel, S. Wuttke, J. Andreato, O. Yaghi, B. Zhang, C. T. Yavuz, T. S. Nguyen, F. Zamora, C. Montoro, H. Zhou, A. Kirchon and D. Fairen-Jimenez, *Adv. Mater.*, 2022, **34**, 2201502.
- 31 K. Mizrahi Rodriguez, W.-N. Wu, T. Alebrahim, Y. Cao, B. D. Freeman, D. Harrigan, M. Jhalaria, A. Kratochvil, S. Kumar, W. H. Lee, Y. M. Lee, H. Lin, J. M. Richardson, Q. Song, B. Sundell, R. Thür, I. Vankelecom, A. Wang, L. Wang, C. Wiscount and Z. P. Smith, *J. Membr. Sci.*, 2022, **659**, 120746.
- 32 R. O'Hayre, S. Cha, W. Colella and F. B. Prinz, *Fuel Cell Fundamentals*, John Wiley & Sons, Inc., 3rd edn, 2016.
- 33 D. Aaron, Z. Tang, A. B. Papandrew and T. A. Zawodzinski, *J. Appl. Electrochem.*, 2011, **41**, 1175–1182.
- 34 M. Guarnieri, A. Trovò, G. Marini, A. Sutto and P. Alotto, *J. Power Sources*, 2019, **431**, 239–249.
- 35 A. C. Lazanas and M. I. Prodromidis, *ACS Meas. Sci. Au*, 2023, **3**, 162–193.
- 36 J. Schneider, T. Tichter and C. Roth, in *Flow Batteries*, ed. C. Roth, J. Noack, M. Skyllas-Kazacos, John Wiley & Sons, Inc., 2023, pp. 229–262.
- 37 A. Forner-Cuenca, E. E. Penn, A. M. Oliveira and F. R. Brushett, *J. Electrochem. Soc.*, 2019, **166**, A2230.
- 38 D. Reber, J. R. Thurston, M. Becker and M. P. Marshak, *Cell Rep. Phys. Sci.*, 2023, **4**(1), 101215.
- 39 P. L. Domingo, B. Garcia and J. M. Leal, *Can. J. Chem.*, 1987, **65**, 583–589.
- 40 E. M. Fell, D. D. Porcellinis, Y. Jing, V. Gutierrez-Venegas, T. Y. George, R. G. Gordon, S. Granados-Focil and M. J. Aziz, *J. Electrochem. Soc.*, 2023, **170**, 070525.
- 41 R. M. Darling and M. L. Perry, *J. Electrochem. Soc.*, 2014, **161**, A1381.
- 42 M.-A. Goulet and M. J. Aziz, *J. Electrochem. Soc.*, 2018, **165**, A1466.
- 43 H. O'Connor, J. J. Bailey, O. M. Istrate, P. A. A. Klusener, R. Watson, S. Glover, F. Iacoviello, D. J. L. Brett, P. R. Shearing and P. Nockemann, *Sustainable Energy Fuels*, 2022, **6**, 1529–1540.
- 44 M. Murbach, B. Gerwe, N. Dawson-Elli and L. Tsui, *J. Open Source Software*, 2020, **5**, 2349.
- 45 A. A. Wong and M. J. Aziz, *J. Electrochem. Soc.*, 2020, **167**, 110542.
- 46 J. D. Milshtein, J. L. Barton, R. M. Darling and F. R. Brushett, *J. Power Sources*, 2016, **327**, 151–159.
- 47 X.-Z. Yuan, C. Song, H. Wang and J. Zhang, in *Electrochemical Impedance Spectroscopy in PEM Fuel Cells: Fundamentals and Applications*, Springer, 2010, pp. 139–192.
- 48 G. J. Brug, A. L. G. van den Eeden, M. Sluyters-Rehbach and J. H. Sluyters, *J. Electroanal. Chem. Interfacial Electrochem.*, 1984, **176**, 275–295.
- 49 A. C. Lazanas and M. I. Prodromidis, *ACS Meas. Sci. Au*, 2023, **3**, 162–193.
- 50 R. M. Darling and M. L. Perry, *ECS Trans.*, 2013, **53**, 31.
- 51 Y. Ji, M.-A. Goulet, D. A. Pollack, D. G. Kwabi, S. Jin, D. De Porcellinis, E. F. Kerr, R. G. Gordon and M. J. Aziz, *Adv. Energy Mater.*, 2019, **9**, 1900039.
- 52 J. Luo, B. Hu, C. Debruler, Y. Bi, Y. Zhao, B. Yuan, M. Hu, W. Wu and T. L. Liu, *Joule*, 2019, **3**, 149–163.
- 53 G. W. A. Foster, *J. Chem. Soc., Trans.*, 1906, **89**, 912–920.



- 54 I. M. Kolthoff and E. A. Pearson, *Ind. Eng. Chem., Anal. Ed.*, 1931, **3**, 381–382.
- 55 R. B. Loftfield and E. Swift, *J. Am. Chem. Soc.*, 1938, **60**, 3083–3084.
- 56 M. Shirom and G. Stein, *J. Chem. Phys.*, 1971, **55**, 3379–3382.
- 57 S. Ašperger, *Trans. Faraday Soc.*, 1952, **48**, 617–624.
- 58 C. A. P. Arellano and S. S. Martínez, *Sol. Energy Mater. Sol. Cells*, 2010, **94**, 327–332.
- 59 M. Reinhard, T. J. Penfold, F. A. Lima, J. Rittmann, M. H. Rittmann-Frank, R. Abela, I. Tavernelli, U. Rothlisberger, C. J. Milne and M. Chergui, *Struct. Dyn.*, 2014, **1**, 024901.
- 60 P. A. Prieto-Díaz, A. A. Maurice and M. Vera, *Chem. Eng. J.*, 2025, **525**, 170162.
- 61 P. A. Prieto-Díaz, A. Trovò, G. Marini, M. Rugna, M. Vera and M. Guarnieri, *Chem. Eng. J.*, 2024, **492**, 152137.
- 62 R. M. Darling, A. Z. Weber, M. C. Tucker and M. L. Perry, *J. Electrochem. Soc.*, 2016, **163**, A5014–A5022.

